Understanding Factors Influencing the Performance of a Wi-Fi Fingerprinting System

Dissertation for the acquisition of the academic degree

Doktor der Ingenieurwissenschaften (Dr.-Ing.)

Submitted to the

Faculty of Electrical Engineering / Computer Science,

University of Kassel

By

M.Sc. Ngoc Doan Duong

Day of Defence: 25th March 2019

Abstract

Location information plays a vital role in today's society. People usually carry their mobile devices everywhere they go to benefit from real-time location services; the location of the device is the location of users. The focus of positioning services is shifting from outdoors to indoors. Technological services which depend on indoor locations are increasing in popularity. Wi-Fi fingerprinting is a promising technique that can be used for indoor localization. In this regard, this dissertation targets at improving understanding on the influence of factors on Wi-Fi received signal strength. It provides useful information applicable in the implementation of a reliable, consistent Wi-Fi fingerprinting system that takes into account factors such as accuracy, recognition rate, and energy consumption.

Different techniques and algorithms have been used in developing a Wi-Fi fingerprinting system. Several studies have been done to analyze factors that influence the performance of a Wi-Fi fingerprinting system. New technologies in wireless networks may provide useful features to improve the performance of Wi-Fi fingerprinting systems but may also give rise to new challenges. Hence, despite the intense research on the field, there are still factors which influence the Wi-Fi signal and performance of Wi-Fi fingerprinting that have not been thoroughly investigated.

In this Ph.D. thesis, I performed various experiments to investigate factors influencing signal strength of a Wi-Fi network and the performance of a Wi-Fi fingerprinting system. I compared the fluctuation of 2.4 and 5 GHz bands by considering factors such as how the presence of people in office environments such as corridors, halls, and office rooms affects Wi-Fi signals. The performance of a Wi-Fi fingerprinting system using the 2.4 and 5 GHz Wi-Fi signal is also evaluated in terms of accuracy, recognition rate, and power consumption in scanning those networks. The influence of small-scale fading and the device heterogeneity problem on Wi-Fi signal strength and Wi-Fi fingerprinting was also be investigated in this thesis. The statistical ANOVA and t-test were used to validate the influence of small-scale fading and device heterogeneity on Wi-Fi signal strength. I analyzed the distribution and the fluctuation of measured Wi-Fi data and then compared the performance of the Wi-Fi fingerprinting system WHERE under the influence of those factors. Consequently, the results showed that the Wi-Fi fingerprinting system achieves similar accuracy when using 2.4 GHz

and 5 GHz bands. However, the recognition rate of a system using signals of 5 GHz was found to be higher than that using 2.4 GHz signals. Scanning 2.4 GHz networks consumes less power than scanning 5 GHz networks. The statistical tests also showed that there is a difference between mean values of Wi-Fi signals measured over a short distance. The Wi-Fi signal strength measured at the same location by different devices is also different. The recognition rate decreases from 100% to 47.76% when heterogeneous devices are used in the training phase and the positioning phase. In addition to device heterogeneity, small-scale fading was also found to impact fingerprints of the measured positions in such a way that devices that were only one centimeter apart were erroneously recorded as different locations. To mitigate the influence of small-scale fading, the collection of Wi-Fi data collected over a small distance can be used to generate the fingerprint of the location and results in an improvement in the recognition to 92.13%.

The results of this Ph.D. thesis help to better understand the different characteristics of the 2.4 and 5 GHz Wi-Fi signals as well as the influence of different factors on the performance of a Wi-Fi fingerprinting system. The selection of frequency bands in Wi-Fi fingerprinting approaches may not influence the results of accuracy but may influence the recognition rate and the power consumption of the system. In this regard, a trade-off of the performance should be considered when designing an indoor localization system using Wi-Fi fingerprinting. I propose to record the motion state of measurement devices when training data is collected. The justification is that when the measurement devices are slightly moved, the collected data was more reliable than when the measurement devices are kept stationary. These understandings provide useful information for the design and implementation of Wi-Fi fingerprinting systems.

Zusammenfassung

Standortinformationen spielen in der heutigen Gesellschaft eine entscheidende Rolle. Menschen tragen ihre mobilen Geräte in der Regel überall hin mit sich, um von Echtzeit-Ortungsdiensten zu profitieren. Hierbei repräsentiert der Standort des Geräts den Standort des Benutzers. Der Fokus von Positionierungsdiensten verlagert sich mehr und mehr von Outdoor-Lokalisation zur Indoor-Lokalisation. Technologische Dienstleistungen auf Basis von Indoor-Lokalisation werden dabei immer beliebter. Wi-Fi-Fingerprinting ist ein vielversprechender Ansatz, welcher für Indoor-Lokalisation verwendet werden kann. Ziel dieser Dissertation ist es, das Verständnis über verschiedene Einflussfaktoren auf die Wi-Fi Signalstärke zu erhöhen. Diese Dissertation liefert nützliche Informationen für die Implementierung eines zuverlässigen, konsistenten Wi-Fi-Fingerprinting Systems, wobei Faktoren wie Genauigkeit, Erkennungsrate und Energieverbrauch berücksichtigt werden.

Bei der Entwicklung von Wi-Fi-Fingerprinting Systemen wurden verschiedene Techniken und Algorithmen verwendet. Verschiedene Publikationen haben Faktoren untersucht, welche die Leistung eines Wi-Fi-Fingerprinting Systems beeinflussen. Neue Standards für drahtlose Netzwerke bieten einerseits nützliche Funktionen, um die Leistung eines Wi-Fi-Fingerprinting Systems zu verbessern, stellen aber auch neue Herausforderungen dar. Trotz intensiver Forschung in diesem Bereich existieren weiterhin Faktoren, welche das Wi-Fi-Signal und die Leistung des Wi-Fi-Fingerprinting Systems beeinflussen aber noch nicht vollständig untersucht wurden.

In dieser Doktorarbeit habe ich verschiedene Experimente durchgeführt, um verschiedene Faktoren zu untersuchen, welche Einfluss auf die Signalstärke eines Wi-Fi-Netzwerks sowie auf die Leistung eines Wi-Fi-Fingerprinting Systems haben. Dafür verglich ich die Fluktuation von 2,4 und 5 GHz-Bändern unter Berücksichtigung wie die Anwesenheit von Menschen in Arbeitsumgebungen, wie zum Beispiel Flure, Hallen oder Büroräume, Wi-Fi Signale beeinflussen. Weiterhin wurde die Leistung eines Wi-Fi-Fingerprinting Systems mit 2,4 und 5 GHz Wi-Fi-Signalen in Bezug auf Genauigkeit, Erkennungsrate und Stromverbrauch beim Scannen von Netzwerken evaluiert. Der Einfluss von Small-Scale Fading und Geräteheterogenität auf die Wi-Fi-Signalstärke und das Wi-Fi-Fingerprinting wurde in dieser Arbeit ebenfalls untersucht. Um den Einfluss von Small-Scale Fading und der Geräteheterogenität auf die Wi-Fi-Signalstärke zu validieren

wurden Varianzanalysen und t-Tests verwendet. Ich analysierte die Verteilung und die Fluktuation der gemessenen Wi-Fi-Daten und verglich daraufhin die Leistung des Wi-Fi-Fingerprinting Systems WHERE unter dem Einfluss dieser Faktoren. Die Ergebnisse zeigen, dass das Wi-Fi-Fingerprinting System eine ähnliche Genauigkeit erreicht, wenn 2,4 GHz und 5 GHz Bänder verwendet werden. Die Erkennungsrate eines Systems mit 5 GHz Signalen war jedoch höher als ein System mit 2,4 GHz. Außerdem verbraucht das Scannen von 2,4-GHz-Netzwerken weniger Energie als das Scannen von 5-GHz-Netzwerken. Die statistischen Auswertungen zeigen ferner, dass der Mittelwert der Wi-Fi-Signale, welche über verschiedene, kurze Distanzen gemessen wurde, variiert. Die am gleichen Ort von verschiedenen Geräten gemessene Wi-Fi-Signalstärke ist ebenfalls unterschiedlich. Die Erkennungsrate sinkt von 100% auf 47.76%, wenn heterogene Geräte in der Trainingsphase und der Positionierungsphase eingesetzt werden. Neben der Geräteheterogenität hatte auch Small-Scale Fading einen Einfluss die Fingerabdrücke der gemessenen Positionen in der Art, dass Geräte, welche lediglich wenige Zentimeter voneinander entfernt waren, (fälschlicherweise) als unterschiedliche Positionen betrachtet wurden. Um den Einfluss von Small-Scale Fading zu minimieren, kann die Sammlung von Wi-Fi-Daten, die über eine kleine Entfernung gesammelt werden, verwendet werden, um den Fingerabdruck des Standorts zu erzeugen, durch welche die Erkennungsrate auf 92,13% verbessert wird.

Die Ergebnisse dieser Doktorarbeit helfen, die unterschiedlichen Eigenschaften der 2,4 und 5 GHz Wi-Fi-Signale sowie den Einfluss verschiedener Faktoren auf die Leistung eines Wi-Fi-Fingerprinting Systems besser zu verstehen. Die Auswahl der Frequenzbänder in Wi-Fi-Fingerprinting-Ansätzen beeinflusst nicht notwendigerweise die Genauigkeit, kann jedoch Einfluss auf die Erkennungsrate und den Stromverbrauch haben. Der Trade-Off zwischen Genauigkeit, Erkennungsrate und Energieverbrauch sollte bei der Entwicklung eines Indoor-Lokalisierungssystems, welches Wi-Fi-Fingerprinting benutzt, berücksichtigt werden. Ich empfehle daher, den Bewegungszustand der Messgeräte bei der Erfassung der Trainingsdaten mitaufzuzeichnen. Sofern sich die Messgeräte geringfügig bewegt wurden, waren die aufgezeichneten Daten verlässlicher, als wenn die Messgeräte sich in einem unbewegten Zustand befanden. Diese Erkenntnisse stellen nützliche Informationen für die Entwicklung und Implementierung von Wi-Fi-Fingerprinting Systemen dar.

Acknowledgment

This thesis has a single author, but it is the result of years of collaboration. To finish my Ph.D., first and foremost, I would like to thank my supervisor and mentor Prof. Dr.-Ing Klaus David, head of the Chair for Communication Technology (ComTec), University of Kassel, Germany. He offers me an excellent chance to do research in ComTec, gets me interested in the topic and guides me with valuable advice on my research.

I would like to thank Prof. Dr. sc. techn. Dirk Dahlhaus, Prof. Dr.-Ing. Axel Bangert, and Prof. Dr.-Ing. M. Eng. Dieter Wloka for their reviews of my Ph.D. I am grateful for my colleagues, Dr. Yaqian Xu, and all members in ComTec who have supported and inspired me to over difficulty of the research works. Moreover, they also give me useful help and advice for my studying and life in a foreign country.

Also, I would like to express my thank to my friends in Germany and in Vietnam, who accompany me during my whole Ph.D. life or a session. I would like to show my deepest thanks to my wife, my son, and my parents, who always stay with me, help, encourage me and share with me the sadness and joyfulness moment.

I owe a great deal to other scholars cited in this thesis. Their works and the insights of their publications give me useful knowledge, information to finish my thesis.

I also thank very much for organizations that give me the financial support for my Ph.D. study: Ministry of Education and Training of Vietnam, ComTec, DAAD scholarship and support programme (STIBET), and the University of Kassel.

Without the help of those people and organizations, I cannot finish my study. Thank you very much.

Contents

Abbreviations

The list below gives an overview of abbreviations used throughout the thesis.

ANOVA	Analysis of Variance
AOA	Angle of arrival
AP	Access point
BSSID	Basic service set identifier
CDF	Cummulative density function
CSI	Channel state information
CSMA/CA	Carrier Sense Multiple Access / Collision Avoidance
DSSS	Direct Sequence Spread Spectrum
EIRP	Equivalent Isotropically Radiated Power
ETSI	European Telecommunications Standards Institute
FCC	Federal Communications Commission
FHSS	Frequency Hoping Spread Spectrum
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
HLF	Hyperbolic Location Fingerprinting
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
LAN	Local area network

LoS	Line of sight
MAC	Media access control
MIMO	Multiple input multiple output
Non-LoS	Non line of sight
OFDM	Orthogonal Frequency Division Multiplexing
QAM	Quadrature amplitude modulation
RFID	Radio frequency identification
RSS	Received signal strength
RTOF	roundtrip time of flight
SMN	Spatial mean normalization
SSID	Service set identifier
TDOA	Time difference of arrival
TOA	Time of arrival

1 Introduction

A context-aware system can support people in various kinds of services and applications. For example, the ability of an application to recognize a user's physical activity can be used to assist elderly people in their daily living. Context awareness is an enabling technology for an autonomous computing system. Location awareness is one kind of context. Today, location information plays a vital role and can be utilized in many applications and services. The focus of the positioning services has changed from mostly outdoors to indoors [1]. Currently, people tend to share more location information with others, especially over their social networks. As the number of mobile phone users and social network users increase rapidly, sharing of location information is also bound to increase [2]. Moreover, people usually carry their mobile devices everywhere they go; the location of the device is the location of users. Hence, smartphone applications take advantage of the knowledge of their location to provide users with better services. For example, a context-aware application may recognize that the person is in the living room and turn on the light automatically.

Wireless LANs (Wi-Fi) are becoming ubiquitous. The wireless LANs infrastructures are deployed in multiple areas such as public places, office buildings, commercial centres, airport lounges, hotel meeting rooms, cafeteria, and private households across the globe. Thanks to the extensive deployment of Wi-Fi, Wi-Fi fingerprinting has emerged as an approach that is suitable for indoor positioning. Wi-Fi fingerprinting utilizes a Wi-Fi pattern from available Wi-Fi access points (APs) to locate the position of user/device indoor. In this thesis, I investigate factors that influence Wi-Fi signals and the performance of Wi-Fi fingerprinting systems in their implementation.

1.1 Problem Statements

Wi-Fi fingerprinting systems leverage on Wi-Fi received signal strength (RSS) from surrounding Wi-Fi access points to generate Wi-Fi fingerprints and locate user positions. The critical assumption made in the use of Wi-Fi fingerprinting is that the Wi-Fi signal strength does not vary over time and the fingerprint is unique for each location. However, the Wi-Fi signal strength measured from an access point changes over time due to various causes. For instance, changes of the environment such as the presence of people could cause variation of the signal. Additionally, the movement of measurement devices over very short distances may experience small-scale fading which result in severe fluctuation of RSS. Another challenge for the Wi-Fi fingerprinting system is the influence of device heterogeneity problems on the performance of fingerprinting systems. Using different hardware for the training and testing phase may result in different received signal strength values and also affect the performance of the system. The distribution of Wi-Fi RSS values, their temporal variation may influence the performance of the Wi-Fi fingerprinting systems. These factors lead to a challenge of how to collect Wi-Fi RSS efficiently and subsequent building distinctive location fingerprints to achieve high localization accuracy.

Frequency is another factor that may influence W-Fi signal strength. Wi-Fi networks operate on both 2.4 GHz and 5 GHz bands. The use of different frequency bands may result in different RSS value and cause the different performance of a Wi-Fi fingerprinting system. Therefore, it is necessary to investigate the characteristics of the 2.4 GHz and 5 GHz Wi-Fi signals, compare the performance of a Wi-Fi fingerprinting system using 2.4 and 5 GHz signals. Understanding the influence of the factors on Wi-Fi RSS provides useful information for implementing a reliable, consistent Wi-Fi fingerprinting system that considers the accuracy, recognition rate, and energy consumption.

1.2 Contributions of the thesis

In this thesis, I investigate factors that influence Wi-Fi signal strength and the performance of Wi-Fi fingerprinting systems. Based on the results, I suggest the selection of one of the two 2.4 or 5 GHz frequency bands in implementing a Wi-Fi fingerprinting system regarding the result of accuracy, recognition rate, and power consumption. I also outlines recommendations for collecting Wi-Fi signals in the training phase to improve the performance of the system.

First, the fluctuation of Wi-Fi signal in office environments such as halls, corridors and rooms was investigated while considering how the presence of people influences signal strength. The fluctuation, the signal distribution, and the fingerprint range of 2.4 GHz and 5 GHz networks, regarding the scenario of indoor space with several small rooms divided by walls, are analysed. In addition to the analysis of the signals, the accuracy and recognition rate of the Wi-Fi fingerprinting system WHERE [3], [4] using these different frequency bands was analysed. Furthermore, the power consumption of the scanning process in the 2.4 GHz and 5 GHz networks was compared and analysed.

In addition, the fluctuation of Wi-Fi signals by the influence of small-scale fading and device heterogeneity was experimentally studied and validated by the statistical Analysis of Variance (ANOVA) test and t-test. Then, these influences are examined carefully by comparing the recognition rate of a Wi-Fi fingerprinting application under such influences. Besides, I also compare the performance of Wi-Fi fingerprinting in the experimental scenarios and the real scenarios with the consideration of the small-scale fading problem.

The main contribution of this thesis is to provide a better understanding of factors influencing the performance of Wi-Fi fingerprinting systems. The results of this thesis help provide a clearer understanding of the effects of fluctuation of Wi-Fi signals in office environments and the power consumption to perform Wi-Fi scanning task in the 2.4 GHz, 5 GHz, and both band signal. The fingerprint range of 2.4 and 5 GHz signal is different which then influences the performance of Wi-Fi fingerprinting systems. Statistical tests help to prove the influence of small-scale fading and device heterogeneity on the Wi-Fi RSS values; the results show that the accuracy of a Wi-Fi fingerprinting system is degraded under the influence of small-scale fading and the device heterogeneity. Subsequently, a method of mitigating the influence of small-scale fading with the assistance of the embedded accelerometer sensors was proposed.

1.3 Outline of the thesis

This thesis is organised in six chapters. The problem of Wi-Fi fingerprinting and the contribution of the thesis are introduced in the first chapter. In chapter 2, the state of the art is presented where the fundamental positioning techniques, challenges in implementing a fingerprinting system, analysis methods and tools are outlined. In chapter 3, I present the analysis of the fluctuation of Wi-Fi signal in an office environment. In chapter 4, the performance of a Wi-Fi fingerprinting system using 2.4 and 5 GHz signals is compared. In chapter 5, I present an investigation of the influence of small-scale fading and device heterogeneity on Wi-Fi RSS and performance of a fingerprinting system. A conclusion is finally provided in chapter 6.

1.4 Publications

Parts of the work conducted for this thesis have already been published at conferences or workshops. These publications are as follows:

- D. Duong, Y. Xu, and K. David, “The Influence of Fast Fading and Device Heterogeneity on Wi-Fi Fingerprinting,” in Proceedings of IEEE 87th Vehicular Technology Conference (VTC2018-Spring), Porto, Portugal, 4 – 6 June 2018.
- D. Duong, Y. Xu, and K. David, “Comparing the Performance of Wi-Fi Fingerprinting using the 2.4 GHz and 5 GHz Signals,” in Proceedings of IEEE 87th Vehicular Technology Conference (VTC2018-Spring), Porto, Portugal, 4 – 6 June 2018.
- Y. Xu, D. Duong, and K. David, “How Near Is Near: A Case Study of the Minimum Distance to Distinguish Neighbouring Places in Place Learning Using Wi-Fi Signals,” in Proceedings of IEEE 83rd Vehicular Technology Conference (VTC2016-Spring), China, 2016, pp. 1–5.

2 State of the Art

In this chapter, I introduce the state of the art concerning location as a context, positioning applications, positioning methods used for outdoor and indoor purposes, background knowledge for understanding the Wi-Fi fingerprinting system, how Wi-Fi fingerprinting works, previous publications related to Wi-Fi fingerprinting, and different factors that may influence the Wi-Fi signal and Wi-Fi fingerprinting.

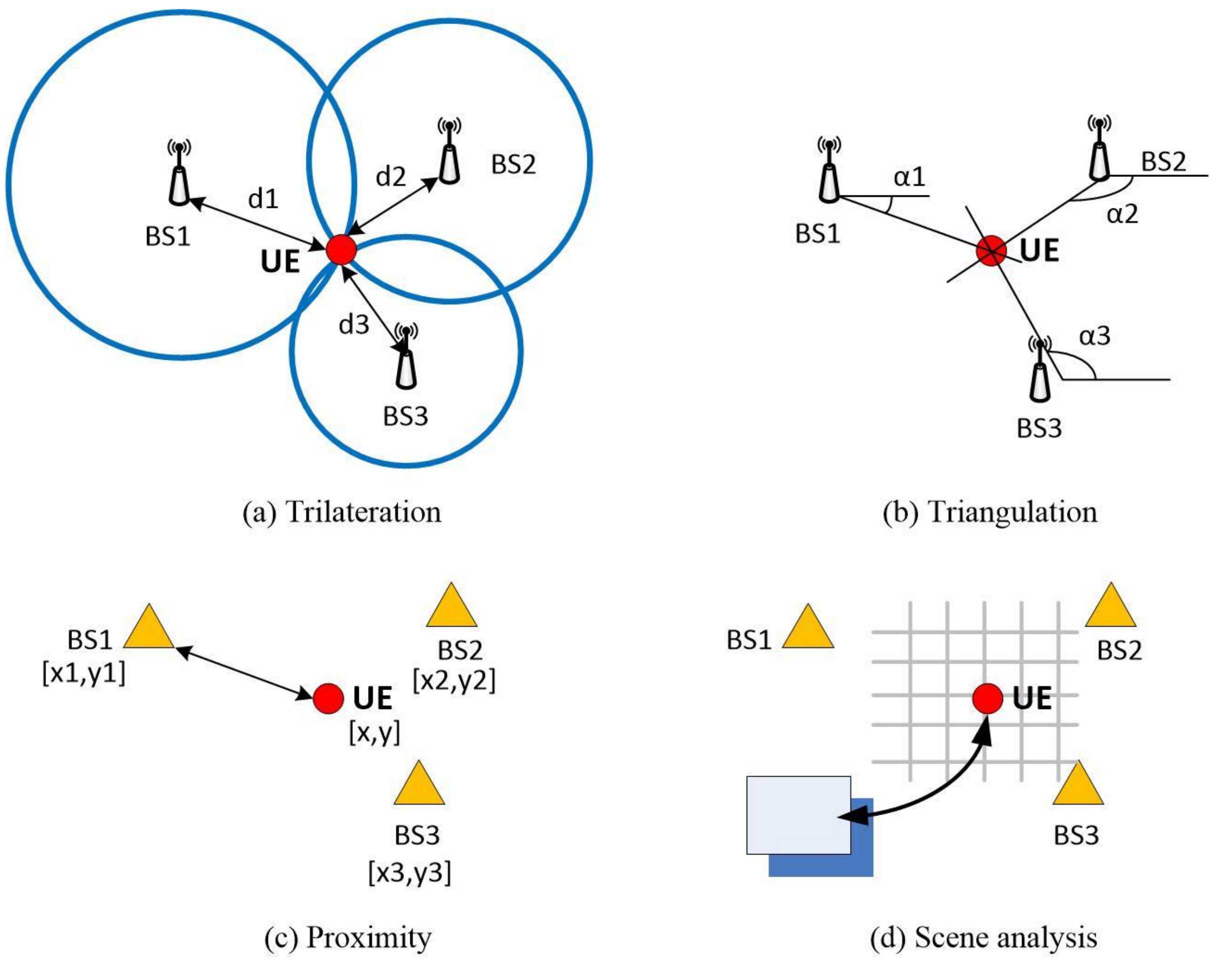
2.1 Positioning applications

Location awareness is a fundamental and essential function for many applications. In daily life, people often perform different kinds of activities at specific locations, so location information is a user's context which indicates people's activities. For instance, knowledge of a user's location is a useful context that can provide effective solutions for work, health, social, entertainment activities, and many more [5]. Therefore, if the positioning applications can get the exact location of users, they may infer their activities and provide services suitable to their immediate context. Thanks to the development of modern technologies, equipment such as mobile phones, sensors, and electronic devices offer a broad range of possibilities to gather information for context recognition and prediction.

The mobile phone is one of the most widely adopted technologies in history and a popular device for everyone. In [2], Frith mentions that besides the traditional function of communication through the use of mobile phones, the smartphone is also used as a locative media. The usage of services using positions on users' devices has shifted from on-demand navigation capabilities to always-on positioning services such as weather updates, travel information, location-based reminders, and so on. Since people carry their smartphones or mobile devices wherever they go, the location information obtained from smartphones provides useful context to reflect user's activities in their daily living. Among various applications, users use their smartphones for services provided by applications related to positioning frequently. The mobile map application is one of the most popular location-based services offered by mobile phones. People use mobile maps to know their location, track their routes, get the accurate direction guidance and navigation assistance from the departure to the destination location. People also frequently use their phones to get information about their surrounding spaces for purposes such as checking in at popular and interesting places or map friends.

The location information can also be used to guide people in unfamiliar buildings such as for guiding passengers to specific gates in the airport, train station, or assisting visitors in a museum. It can also be used to track objects such as books in a library, or assets in a warehouse. For entertainment, interactive games can also benefit from the indoor location by tracking the location of the body parts of a player and making adjustments to enhance the experience [6]. Another application is video or audio playback applications that may track the current location of users to automatically turn the system on or off [7]. For advertising, customer location can be utilized to provide targeted advertisements of product information inside retail stores [8]. In general, the location context information can be utilized to improve people's quality of life.

2.2 Fundamental Positioning techniques



Legend: UE: user equipment; BS: base station; d: distance

Figure 2.1 Fundamental positioning techniques using radio signals [1].

Different kinds of wireless technologies have been leveraged for positioning purpose, including infrared, ultrasound, Bluetooh, WLAN, RFID, magnetic field, etc. Each technology has its advantages and disadvantages in performing the localization activity. Regardless of the positioning technology, different positioning techniques can be used to determine the location of people or objects. They are trilateration, triangulation, scene analysis, proximity, and hybrid methods. The positioning system can combine different positioning techniques as a hybrid solution to increase the performance of the system. Fig. 2.1 shows the fundamental positioning techniques. A positioning system can use one or combine multiple positioning techniques to take advantage of each technique.

- The trilateration technique calculates the distance from the measurement device to several reference points to estimate the location of the measurement device using geometry of circles. Trilateration technique may use the time of arrival (TOA) of a signal, time difference of arrival (TDOA) of the signal from multiple APs, or roundtrip time of flight (RTOF) of the signal to calculate the distance based on the velocity of the radio signal and the travel time. To locate the position of an object, trilateration techniques require signals from at least three reference points; the clocks of the transmitters and receivers must be synchronized, and it needs line of sight path between the transmitters and receivers.
- The attenuation of the signal can also be applied with the trilateration technique to calculate the position of a device. This method based on the principle of signal attenuation during transmission to calculate the traveled distance. The path loss propagation model is used to interpret the received signal strength to physical distance. Then, the distance from at least three reference points can be used to figure out the relative distance to the known location. This method requires a precise model to describe the path loss index. However, generating an accurate model to convert signal strength to distance is not easy [9].
- The triangulation technique is based on the angle of arrival (AOA) of a signal to estimate the position of a device. To calculate the angle correctly and then determine the position, this technique requires the calibration of the antenna array.
- The proximity technique uses a dense grid of sensors installed at reference points to estimate the position of users. When a mobile device is detected by a sensor, the mobile device is considered to be in the location area of that sensor. Different kinds

of sensor such as radio frequency identification (RFID) and infrared sensors … can be used together in the proximity technique.

- The scene analysis technique uses features associated with physical locations to distinguish one location from another. The scene analysis needs to collect features of a scene first which can be considered as fingerprints of the scene. Then, the location of the user's device is estimated by comparing the current measurement with the fingerprint in the database [10]. Both radio and non-radio signals such as Wi-Fi signals, Bluetooth signals, magnetic fields can be used as scene analysis features. Among them, Wi-Fi RSS-based positioning or Wi-Fi fingerprinting is commonly used. This technique requires surveying the area to generate the fingerprint database in advance and a stable environment to have good performance.
- Dead reckoning method uses the inertial sensors of the mobile device such as the accelerometer sensor, gyroscope sensor, and compass sensor to track the path of the target device. The number of steps the person has walked is counted and used to infer the distance. Meanwhile, the direction after each step is also calculated. These data sets are combined to estimate the distance and direction the user has passed and figure out the relative positions compared to the reference point. Dead reckoning method requires knowing the layout of the building so that the location can be mapped. This method can avoid the problem of multipath signal faced in other approaches. However, the challenge of dead reckoning is that the sensors in mobile devices need to be calibrated well to avoid error. Otherwise, the inaccurate number of steps, step length, as well as the direction of walking can lead to the huge error of walking after a period of time. Moreover, the compass sensor data may be influenced by magnetic material inside the building and the step length of each user is not usually the same. Thus, the assumption of the equal step length can lead to errors in calculating the traveled distance.

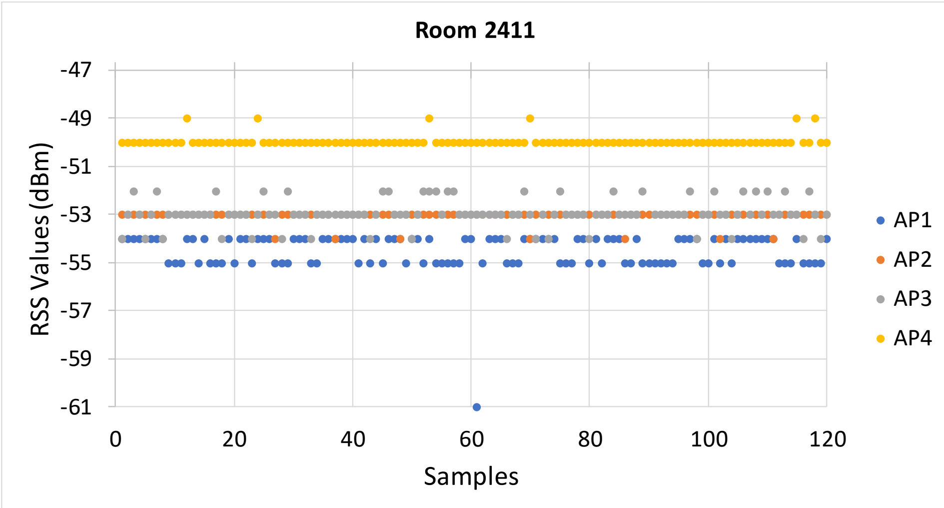
2.3 Wi-Fi fingerprinting approach

Global positioning system (GPS) is a satellite navigation system which is commonly used for the outdoor positioning applications [11]. However, this system is not suitable for the indoor positioning purpose. Assisted-GPS techniques may have an error of tens of meters for indoor positioning [12]. The positioning accuracy requirement for the indoor is higher than that for outdoors [13]. A few meters of accuracy for indoor localization may places people to another office within a building. Indoor localization application can provide benefits for many activities such as entertainment activities, monitoring elderly people, patients, monitoring things in warehouses, environment control, smart home, etc. There is a strong need for a precise, reliable, and quick response of localization in an indoor environment. To extend the capability of localization, other techniques have been developed to substitute GPS for the indoor localization. Different wireless technologies have been used for indoor localization purpose including 802.11 wireless (Wi-Fi), infrared, ultrasonic signal, etc. [14]–[17]. Bat and Cricket [17] combine radio frequency and ultrasound signal to locate the position of users indoor with high accuracy. The system consists of wireless transmitters, receivers, and a central radio frequency (RF) base station. The central RF base station periodically broadcast RF message. When hearing a message, the transmitter sends out an ultrasonic pulse. The receivers receive both the RF signal and the ultrasonic signal and determine the time interval between those two signals to estimate the distance to the transmitter. The Active Badge system [15] uses infrared signal to track objects or users. A badge worn by users periodically transmits its unique identification (ID) using an infrared transmitter. Receiver sensors placed at fixed locations receive the information from the badge to identify the location. However, those systems require to install a large number of sensors, and the transmission range of these sensors is limited. Ultrasound and camera systems provide people a satisfactory accuracy, but they require lots of human effort and money to deploy infrastructure.

Other positioning systems leverage the Wi-Fi network to locate the position of users in an indoor environment. 802.11 wireless technology has developed considerably in recent years to become a ubiquitous wireless network in homes, offices, and public areas. Using existing infrastructure with no specialized hardware required for positioning is a very attractive option because this helps to save time and money in the implementation of an indoor localization system. Therefore, Wi-Fi networks have been utilized for positioning purpose in indoor environments. One method of using Wi-Fi signal for positioning is to convert the Wi-Fi received signal strength to distance measurements by applying the trilateration technique [10]. However, it is difficult to generate an accurate model to convert signal strength to distance because of the complicated propagation of radio signals in an indoor environment [18]. The trilateration or triangulation require line-of-sight between the transmitter and the receiver. Thus, those schemes do not work well in an indoor environment with obstacles and room partitions.

In recent years, another method called Wi-Fi fingerprinting has been actively studied and becomes a promising approach for indoor localization. Wi-Fi fingerprinting has demonstrated the good performance for indoor positioning due to its cost advantage, widely deployed, large coverage indoor and localization accuracy [11]. The principle idea of the Wi-Fi fingerprinting approach is using specific received signal strength pattern of neighboring wireless LAN APs to distinguish different locations. In other words, the Wi-Fi signal measured from surrounding APs at a particular location is used as the fingerprints represented this specific location. The strong point of this approach is that it does not require either to know the exact location of APs or to perform the distance or angle measurement. Therefore, Wi-Fi fingerprinting has a high feasibility that supports its implementation in indoor circumstances.

A Wi-Fi fingerprint consists of a set of Wi-Fi MAC addresses and RSS observed during a scanning period of all Wi-Fi channels. This is similar to the way people use a human fingerprint to differentiate and recognize different people. Fig. 2.2 demonstrates the Wi-Fi signal measured at two adjacent rooms. At those two locations, the measurement device can capture the Wi-Fi signal from the same APs, but the signal strength values or pattern of the AP measured at different places are different. Wi-Fi fingerprinting leverages this feature to generate the unique Wi-Fi fingerprint to distinguish different locations. The performance of a Wi-Fi fingerprinting system depends on the quality of the collected signal used to generate Wi-Fi fingerprint, the accuracy of the fingerprinting database, and the positioning algorithms. Subsequent sections will mention these elements in detail.



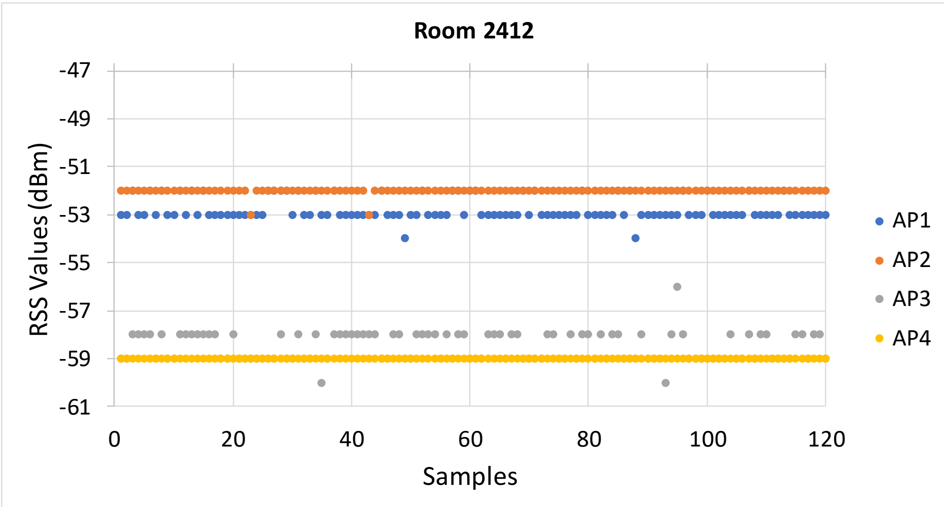


Figure 2.2 Wi-Fi signals measured at two different adjacent rooms.

Wi-Fi fingerprinting localization is divided into two phases of operation: a training phase (also known as offline phase) and a positioning phase (also known as online phase). In the training phase, the received signal strength of Wi-Fi access points is collected at different places covering the area of interest such as the inside a building. This process is also known as the site survey process. Each measurement location is considered as a training point. The Wi-Fi data measured in all training points with their associated names is used to generate the Wi-Fi fingerprint. These fingerprints are stored in a fingerprint database. The Wi-Fi fingerprinting database (also known as a radio map) is a database consisting of pre-recorded measurements of Wi-Fi received signal strength, denoted as location fingerprints. In the positioning phase, the momentary Wi-Fi scan is compared with each of the Wi-Fi fingerprints stored in the fingerprint database to recognize the likeliest Wi-Fi fingerprint and figure out the user's current location

[19]. Fig. 2.3(a) demonstrates the basic fingerprinting system flow including training and positioning phase. Fig. 2.3(b) shows the site survey at several training points with three APs installed at fixed locations. In the training phase, the measurement device was moved to various training points to measure the RSS of those three APs to generate Wi-Fi fingerprints of the area.

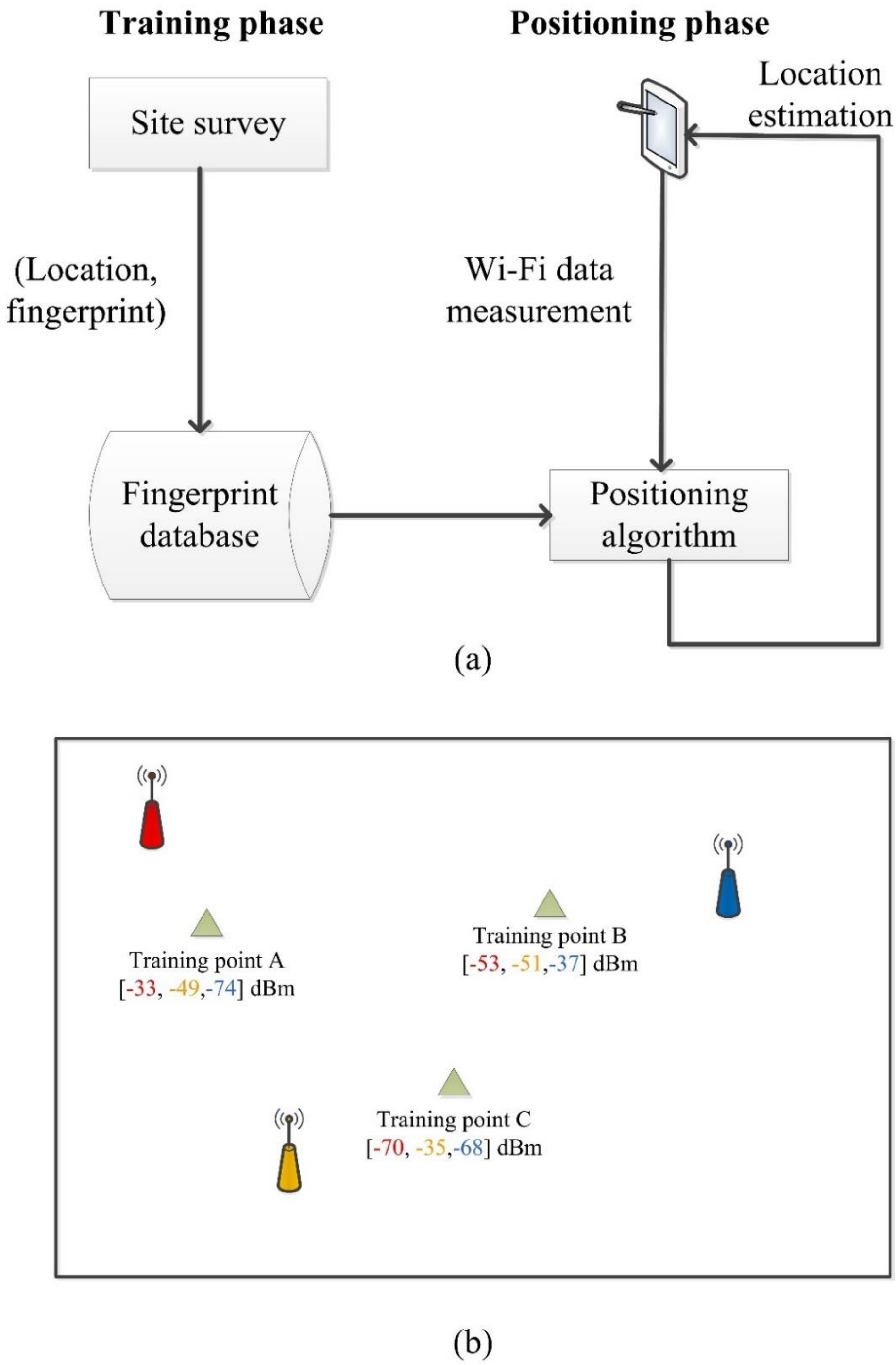


Figure 2.3 The basic fingerprinting system flow and site survey at several training points (based on [20]).

Many localization systems apply fingerprinting technique that discover the signal characteristics (pattern) in certain locations to form fingerprints of these places. Microsoft's RADAR [14] used radio signals to locate and track users inside buildings. In this research, signal strength information at multiple receiver locations was recorded to calculate user's coordinates and infer location when the same characteristics are seen. The BeaconPrint algorithm [21] uses Wi-Fi, and GSM response-rate histograms as a fingerprint to distinguish locations. First, BeaconPrint learned the location's fingerprint based on the stability of the GSM and Wi-Fi information during a pre-defined time window. Then, when the device returns to locations that have been learned, those locations can be recognized by comparing the observed fingerprint with the location's fingerprints. Horus system [22] uses the signal strength distribution information from surrounding APs and probability technique to infer the location of a user. The authors try to reduce the influence of temporal variation by modeling the RSS distribution. Kaemarungsi et al. [23] have listed the factors that may affect the localization performance of a Wi-Fi fingerprinting system. The difference of hardwares, variation of RSS, changes in the environment such as human movement, furniture relocation are factors that cause challenges to the performance of a Wi-Fi fingeprinting system.

Density-based clustering [24] is a well-known and an attractive method to cluster objects with arbitrary shapes and handle the signal noise based on neighborhood density in a given radius (Eps) and a minimum number of points (MinPts). Density-based clustering connects points within a specific radius; if the number of points within a specific radius is higher than the specific MinPts, that group of points is identified as a cluster. A high-density distribution of data may indicate that users spend much of their time at those locations, whereas low-density distribution may determine the non-significant places. ARIEL [5] automatically learned room fingerprints by generating clusters on the collected Wi-Fi scans. The system applies a density-based clustering algorithm to cluster Wi-Fi signals collected in a stationary state to identify zones. Each zone corresponds to one of the stationary occupancy hotspots and is represented by a Wi-Fi signature which consists of a set of Wi-Fi signal vectors. Then, ARIEL measures the motion of users, applies motion based-clustering algorithm to detect inter-zone existing in the same room. Those zones are combined as the room's fingerprint and serve as room identification to distinguish different rooms. In [25], Dousse et al. applied a density-based clustering algorithm OPTICS [26] from raw Wi-Fi measurements to identify significant places based on Wi-Fi signal relative density instead of absolute density threshold. OPTICS enable the use of a sophisticated threshold to detect clusters which appear as local minima. The depth of the local minima depends on the density of the clusters. In this way, this algorithm helps to detect clusters of various densities. The authors conclude that their method can identify an individual's most significant places. The Density-based Clustering Combined Localization algorithm (DCCLA) [4], [27] constructs a fingerprint database from collected Wi-Fi data via mobile phones in users' daily lives, and then separately applies density-based clustering on the RSSs from each APs. Fingerprints of meaningful positions are learned by analyzing the dataset structures of Wi-Fi RSSs.

2.3.1 Scanning Wi-Fi channels

Wireless LANs (WLAN) transmit radio frequency energy through the air. Wi-Fi receiver can pick up radio waves broadcasted on a given frequency if the receiver is tuned to that same frequency. The usable range of Wi-Fi signal depends on transmit power, distance, and interference from other signals or obstacles [12]. To connect to a Wi-Fi access point, mobile devices need to do three following steps in order [28].

- Discover the available APs or scanning the Wi-Fi network
- Authenticate with an AP in which the mobile device wants to connect to
- Associate with that AP

There are specific procedures to perform those above steps to successfully connect a WLAN device to a wireless network. However, the Wi-Fi fingerprinting system does not need to connect to the Wi-Fi network but just needs to get Wi-Fi information from surrounding APs. Therefore, the device only needs to perform the first step: discover the available APs or scan the Wi-Fi network. Scanning the Wi-Fi network is the process of listening to beacon frames broadcasted by surrounding APs to get necessary information about a specific AP. The beacon frame usually consists of a timestamp, MAC address or BSSID, SSID, frequency and current signal strength values of AP. This information is utilized to generate the fingerprints for a Wi-Fi fingerprinting system.

The Institute of Electrical and Electronics Engineers (IEEE) 802.11 standard subdivides the used radio spectrum into a set of channels. In Europe, the 2.4 GHz band has 13 available channels from 2.402 GHz up to 2.480 GHz, while the 5 GHz band has 19 available channels from 5.13 GHz up to 5.805 GHz [29], [30]. The Wi-Fi scanning process tunes from the first to the last channel in the channel list to listen to the beacon frames from nearby APs. Beacons are sent out using the mandatory 802.11 carrier sense multiple access/collision avoidance (CSMA/CA) algorithm with the lowest mandatory data rate. The 2.4 GHz network broadcasts beacons with a bit-rate of 1 Mbit/s using direct sequence spread spectrum (DSSS), whereas the 5 GHz network broadcasts beacons with a bit-rate of 6 Mbit/s using orthogonal frequency division multiplexing (OFDM) [31].

There are two Wi-Fi scanning methods: active scanning and passive scanning.

- In the passive scanning method, the scanning device passively waits to listen to the beacon frames broadcasted from APs. The waiting time in each channel is not defined by the IEEE 802.11 standard. In general, an AP is set to broadcast beacons periodically with an interval of 100ms. A scanning device listens to every channel on the channel list for a given period, then moves to the next channel.
- In the active scanning method, the scanning devices actively request the APs to send the beacon frames to them. The scanning device broadcasts a probe request frame which contains its address and waits for a certain period of time to receive responses from APs. After receiving the probe request frame, APs reply by sending out the probe response frames, which contains similar information as a broadcasted beacon. Then, the scanning device moves to the next channel and repeats the above steps. The process is iterated until all channels have been scanned.

Active scanning does not need to wait for the beacons, so this method helps to save time. On the other hand, passive scanning can reduce workload and save battery power due to its passively listen to APs. To scan Wi-Fi AP, Android platforms currently apply passive scanning as the default method.

2.3.2 Wi-Fi signal collection

The fingerprinting approach requires a survey of an area (site survey) to collect Wi-Fi signals and subsequently generate the fingerprints. However, Wi-Fi signal collection and maintenance are tedious tasks since site survey demand intensive manual labor and time-consuming process to survey a whole area. A grid-based approach is a typical approach to performing site surveys. The survey area is divided into many small grids; at each grid point, the Wi-Fi signal from nearby Wi-Fi APs is scanned for a duration of time or to collect several Wi-Fi samples. Then, the data is used to generate fingerprints representing those locations.

Another method which helps to reduce the time and effort required when collecting Wi-Fi signals is the path survey approach. In this approach, a person carries measurement devices and moves continuously around the surveying area to obtain the Wi-Fi data. The user's walking distance is also recorded during the surveying period. The walked distance will be mapped with the locations, and the fingerprint of locations along the path is generated [32]. In [33], the authors proposed to use the inertial motion sensor to record the walking path and direction when the user is walking. Based on the RSS patterns and these relative distances, direction, the system maps the collected RSS fingerprint to the indoor map. This method helps to collect Wi-Fi signal quickly; users do not need to survey the building intensively. The site survey can be done transparently when users are working with their daily routine. However, the difficulty of this approach is the accumulated distance error after a duration of walking when user's paths are difficult to track.

Interpolation-based is another approach which leverages a signal's propagation model to infer the Wi-Fi signal for the interpolated locations from the observed locations. The whole area is divided into observed and interpolated locations. First, the Wi-Fi signal is measured at observed locations, and the fingerprints of observed locations are generated. The fingerprints of interpolated locations are then interpolated from the observed location fingerprints. Several publications reported different procedures and the different number of observed locations in their studies. In [34], the author proposed scanning four Wi-Fi scans per room, each one in each corner of the room. In [35], Kubota et al. compare location accuracy when selecting a different number of observed locations in all locations.

For indoor positioning purpose, maybe people do not need to know the latitude and longitude of the location where they are, but they want to know if they are in the meaningful places such as living room, bedroom, kitchen room, etc. The meaningful places can be recognized by analyzing users' smartphone data. People often visit and stay for a duration of time in places which are meaningful to them. Based on GPS or accelerometer sensors, a smartphone can detect the mobility and the duration of staying for a specific place. When the smartphones discover the places where people spend more time there, they can automatically learn the fingerprint for these locations in an unsupervised manner without human intervention.

Beaconprint [21] recognizes the significant locations if the users stay in those places for a defined time based on the stability of Wi-Fi and GPS scan.

Other publications present the crowdsourcing approach to reduce time consumption and labor costs [36]–[40]. Crowdsourcing is a method which collects Wi-Fi data from different users carrying a smartphone or laptop. In this method, different users contribute and associate their Wi-Fi data while they are walking around the building and send the data to the fingerprinting system. Users can also report their current locations to the system or let the system figure out the location of users by itself. If the system recognizes the location, it sends the location estimates to users and allows them to correct their location. The location information received from users is stored in the system to aid in the learning of specific fingerprint of places. Crowdsourcing is a cheap and practical method to implement a pervasive Wi-Fi fingerprinting system. However, the device heterogeneity problem which users use different kinds of measurement devices is a major problem of crowdsourcing approach. The device diversity problem may adversely impact the performance of the positioning system [36]. Moreover, require users to give instant feedbacks about their locations may bring them uncomfortable experiences. Additionally, in order to get good quality of data during the collection process, users may need to have some training before they collect the Wi-Fi data.

Different collecting data approaches are summarized in Table 2.1. The major concern in collecting the Wi-Fi data is how to balance the cost, labor effort, and localization accuracy. The traditional grid-based approach is labor intensive but provides real and high-quality data. Other approaches help to save on costs and labor but pose other challenges. Therefore, professional intensive site survey may be needed for an area with high accuracy demand; for areas with low accuracy demand, cost-effective approaches can be applied to reduce cost and labor.

Table 2.1 Wi-Fi signal collecting approaches

Survey Approaches	**Description**	**Limitations**
Grid-based	Manually collect Wi-Fi data at many reference locations. Achieve accurate Wi-Fi radio map. [14]	Intensive site survey, labor and time consuming
Path survey	Continuously move along the survey area. The walking distance and directions are measured while collecting Wi-Fi data. Based on the distance and direction, mapping the RSS pattern to the indoor map. Save time and effort to build a fingerprint database, do not need to perform site survey intensively. [32]	Users’ walking distances are difficult to track. It may lead to the accumulation of error of walking distances after a period of time.
Crowdsourcing	Many users contribute their current locations and associated Wi-Fi data while they are walking around. Quickly collect large data from many users. [36]	Depends on the quality of users’ feedbacks; faces the problem of heterogeneous devices.
Interpolation	First, manually collect Wi-Fi data in observed locations; then estimate Wi-Fi data for other interpolated locations. Reduce the time and effort needed to collect Wi-Fi data for the whole area. [35]	Decrease the positioning accuracy. The interpolated data may be not accurate.
Meaningful place learning	Use mobile sensors to discover meaningful places. The Wi-Fi data in those places is collected automatically. Save time and effort; do not need an intensive survey. [5]	Can only collect data in meaningful locations where the user visits frequently.

2.3.3 Features to use as a fingerprint

In Wi-Fi fingerprinting, the use of features as a fingerprint of locations may influence the performance of the system in term of system accuracy, precision, complexity, etc. Different features related to Wi-Fi signal strength have been used as fingerprints of locations such as the mean RSS value, set of Wi-Fi scans, the signal strength ratio of two APs, the difference in Wi-Fi RSS, RSS range, etc. The mean RSS value is one of the common fingerprint features used in many studies. A set of RSS values from surrounding APs is measured for a duration of time. Then, the average RSS value measured from each AP will be calculated and used as the fingerprint of the location. This kind of feature is quite easy to generate but according to some publication [31], [37], it negatively affects the accuracy of the system in comparison to others.

Another fingerprint feature consists of a set of Wi-Fi scans sensed at a location. This feature does not use the average value but uses the whole set of RSS values as a fingerprint feature. In [4], the authors use the RSS-range which has a high density of RSS value measured at a location as fingerprint feature. The histogram of RSS values is also used as a fingerprint to deal with the fluctuation of the signal at the measured location [41]. In this case, a location fingerprint consists of a set of RSS histograms of APs around the measured location. In [31], Farshad et al. compare the accuracy of a Wi-Fi fingerprinting system using seven different fingerprint features including RSS value, the variation of RSS, the most stable subset of APs (stability), how often different APs are seen (constancy) and the subset of APs that are most widely seen across all cells (coverage), hybrid feature constancy + RSS, and constancy + stability.

The received signal strength value depends on the hardware of the measurement device. Several studies use features which do not depend on the receiver hardware as location's FP. Hossain et al. propose using the signal strength difference between pairs of APs as Wi-Fi fingerprint feature to mitigate the problem of the heterogeneous device [42]. Dong et al. [43] proposed to use the difference between signal strengths across access points as a localization feature. The authors pick out one AP as a reference and subtract its signal strength with the RSS of the other APs to form a new feature which is then used as a fingerprint. They reported that by subtracting the signal strength measured from two APs, the influence of the constant factor of antenna gain is eliminated. Kjærgaard et al. [44] used Hyperbolic Location Fingerprinting (HLF) method which used the signal strength ratios between pairs of base stations as fingerprint feature.

The idea of HLF comes from hyperbolic positioning method, which estimates position from time difference measurements. The author concluded that using signal strength ratios between pairs of base stations is more stable than using absolute signal strength to generate fingerprints. Wang et al. [45] proposed a spatial mean normalization (SMN) method to address the variation in heterogeneous hardware. The SMN method mitigates the difference of the antenna gain among heterogeneous devices by calculating the difference between the absolute RSS and the spatial mean RSS values of the observed APs. Another approach [46] called rank-based fingerprinting uses the rank of the APs as fingerprints. This method sorted the list of visible APs based on their signal strength and assigned them the rank value. The sorted list is stored as the fingerprint for the measured location. In the positioning phase, the list of momentary Wi-Fi scan APs is also sorted and compare with the fingerprint to calculate the similarity. However, the sorted list of APs may be the same in different locations. For example, if we measure the signal in short distant locations (e.g. 1-2 meter), the order of the AP list in those locations may be no different. Therefore, it is a challenge to select the subset of all APs for order comparison.

2.3.4 Positioning algorithms

After collecting Wi-Fi signal data, the Wi-Fi fingerprinting system generates the Wi-Fi fingerprints of reference locations and stores them in the database. In the positioning phase, the Wi-Fi fingerprinting system compares the current Wi-Fi scan with the Wi-Fi fingerprints in the fingerprint database to figure out the current location of user. Since Wi-Fi fingerprinting was first introduced in the year 2000 [14], different learning and positioning algorithms have been used to improve the performance of the system as well as to deal with the challenges arising in a fingerprinting system.

- Nearest neighbor is a common positioning algorithm used in a Wi-Fi fingerprinting system [31]. This method calculates the distance between the value of current Wi-Fi scan and the value of fingerprints from the database. The fingerprint with its associated location which has the shortest distance would be inferred as the current location of the user. Some popular nearest neighbor metrics are Euclidean distance, Manhattan distance, and Mahalanobis distance [31].
- Probability approach calculates the likelihood of the current Wi-Fi scan to each location candidate in the fingerprint database. The location in which the likelihood is highest decided by the decision rule such as Bayesian rule is inferred as the current location.

Different approaches can be used to estimate the likelihood function including histogram, Gaussian, and Log-Normal distribution [47].

- The neural network method uses a structure including many neurons connected in a particular manner to establish a network. The Wi-Fi signal strengths and their associated locations are used as inputs and targets for the training purpose. The output of the training process is an appropriate weight value for each location. In the positioning phase, the Wi-Fi RSS values of APs are used as input data to feed into the neural network to calculate the probability of the input data to each location. The result which has the highest probability is the estimated location [48].
- Support vector machine (SVM) has been used in location fingerprinting [49], [50]. SVM constructs an optimal hyperplane in high dimensional space to divide vectors of input data into separate classes with the largest distance to the nearest vector of other classes as possible. This kind of hyperplane is called maximum margin hyperplane. The distances between classes are called margin. The vectors closest to the maximal margin hyperplane is called support vector. For positioning, the new data are mapped into the same space to find on which side of the hyperplane the new data fall into. Based on the maximal margin hyperplane and support vector, SVM decides which class the data belong to.

2.4 Factors influencing the Wi-Fi signal

Although Wi-Fi fingerprinting has emerged as a suitable choice for indoor localization, Wi-Fi networks are not designed for the localization purpose. Therefore, using the Wi-Fi network to locate the position of users or devices poses difficult challenges. To maintain the best performance, the Wi-Fi signal and Wi-Fi fingerprints in each location should be unique and should not vary. However, this is not true in reality. There are many factors that may influence on the Wi-Fi RSS as well as the performance of Wi-Fi fingerprinting. To implement a Wi-Fi fingerprinting system, a comprehensive understanding of factors that influence the signal propagation, signal characteristic in a complex environment is necessary and useful.

2.4.1 Different Wi-Fi standards

The IEEE 802.11 wireless LAN standard is defined under IEEE networking standards. 802.11 was ratified in the year 1997 with the speed from 1 to 2 Mbit/s. Since then, wireless LAN has gone through great evolution; the Wi-Fi connectivity has been increased tremendously from 1 Mbit/s to a gigabit with the first release of 802.11ac in 2013. At the current time, there are five major IEEE Wi-Fi standards that have been used popularly: 802.11a, 802.11b, 802.11g, 802.11n, and 802.11ac standard. Fig. 2.4 shows the timeline of 802.11 development [29].

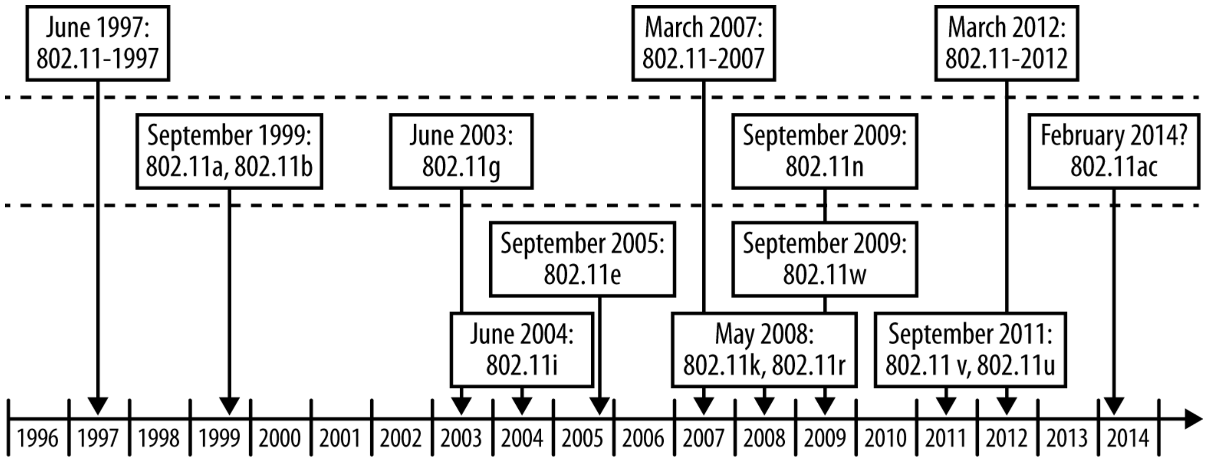


Figure 2.4 Timeline of 802.11 development [29].

Each 802.11 standard supports different data link rates: 802.11b supports 1, 2, 5.5, and 11 Mbit/s; 802.11g was introduced to the market in 2003/2004 and is compatible with 802.11b, provides speeds of up to 54 Mbit/s using OFDM in the 2.4 GHz band. 802.11a was finalized in 1997 in parallel with 802.11b but operates in the 5 GHz band. 802.11a provides high data rate ranging from 6 Mbit/s to 54 Mbit/s; whereas the 802.11ac standard (phase 1) achieves data rate in the range of 1.3 Gbit/s with beamforming and MIMO technique. At the same distance up to 225 feet, the 5 GHz 802.11a throughputs are higher than 2.4 GHz 802.11b systems from 2 times to 4.5 times [12], [51].

Wi-Fi beamforming technique was first introduced in the 802.11n standard which purposed to increase the transmission rate. Beamforming allows the transmitter to beamform its transmitted energy to the receiver with the appropriate phase and amplitude to increase the signal to noise ratio, and hence increase the transmission rate. 802.11ac standard develops this technique to provide higher transmission speed. In 802.11ac, the transmitter and the receiver must exchange information about the characteristics of the channel to set up the explicit beamforming functions. To increase speed, 802.11ac utilizes higher bandwidth per channel up to 160 MHz bandwidth, higher number of spatial streams up to eight, higher order modulation 256 Quadrature amplitude modulation (QAM), and multi-user multiple input and multiple output (MIMO). Before 802.11ac standard, all other 802.11 standards were single user: the transmission was sent to only one receiver at the same time. Multi-user MIMO (MU MIMO) allows sending multiple data frames to multiple users simultaneously. If two receivers located in sufficiently different directions, the beamforming is used to steer each of the transmissions toward its respective receiver using the same channel frequency without causing interference. The multi-user MIMO technology may influence the received signal strength of a receiver if the APs have to share its capability with multiple receivers. Today, APs and mobile devices equipped with Wi-Fi 802.11ac standard and multiple antennas from different manufacturers are now available in the market. Over time, APs and mobile devices will transit from older 802.11 standards to 802.11ac standard. These capabilities enable the increasing the accuracy of Wi-Fi fingerprinting positioning. These new techniques provide consistent performance which addresses the challenge of higher density and more devices connecting to the network [29], [52].

2.4.2 Different frequency bands

Wi-Fi networks operate on both 2.4 and 5 GHz unlicensed bands. Typically, unlicensed bands are rarely interference free because many vendors compete to use the same frequency band for their devices. Different countries or regions of the world have allocated different spectrums for the 2.4 and 5 GHz bands. In Europe, the Wi-Fi 2.4 GHz band is divided into 13 usable channels, and the Wi-Fi 5 GHz band is divided into 19 usable channels. 802.11b and 802.11g operate in the 2.4 GHz band; 802.11n support the dual-band (i.e., 2.4 GHz and 5 GHz), while 802.11a and 802.11ac operate only in the 5 GHz band [29]. Today, the Wi-Fi 2.4 GHz band is heavily used and suffers from the interference of other devices which operate on the same 2.4 GHz band such as Bluetooth devices, microwave oven, etc. Part of the interference problem is caused by the transmit power levels in the 2.4 GHz band. For instance, FCC allows 2.4 GHz Frequency Hopping Spread Spectrum (FHSS), and Direct Sequence Spread Spectrum (DSSS) devices can have a maximum peak output power of 1 W. Thus, a wideband DSSS device can be interfered by a narrow band FHSS device. A narrowband 1 MHz channel Bluetooth device or 2.4 GHz cordless telephones have a high likelihood of hopping into the 22 MHz channel that a 2.4 GHz DSSS Wi-Fi system uses. As the 2.4 GHz band is heavily used, the less crowded 5 GHz band is used to avoid much of the interference at 2.4 GHz. Table 2.2 shows 802.11 standards operating in the 2.4 GHz and 5 GHz bands.

Table 2.2 802.11 standards operating in the 2.4 GHz and 5 GHz bands.

2.4 GHz	**5 GHz**
802.11 (direct sequence and frequency hopping)	802.11a
802.11b	802.11n
802.11g	802.11ac
802.11n	

The coverage range of the Wi-Fi network and its received signal strength depends on the transmit power, the distance between the transmitter and receiver, and the interference of the transmitted environment [12]. Different parts of the 5 GHz spectrum have been set different output power requirements. According to the European Telecommunications Standards Institute (ETSI) standards applying to European countries, for 5 GHz band, the low band operates from 5.15 GHz to 5.35 GHz with a maximum mean Equivalent Isotropically Radiated Power (EIRP) of 200 mW; the middle band operates from 5.47 GHz to 5.725 GHz with a maximum mean EIRP of 1 W; For the 5.725 GHz to 5.875 GHz band, the maximum mean EIRP is 25 mW [12]. The 5 GHz power rules help to mitigate and limit potential interference to 5 GHz wireless LANs. Besides the limited power, the Federal Communications Commission (FCC) has also specified power spectral density limits to force narrower bandwidth systems to transmit with less power. Moreover, 5 GHz unlicensed band is only used for high data rate communications devices. Therefore, 2.4 GHz narrowband interferers such as cordless phones, low rate Bluetooth devices are unlikely to be used in 5 GHz band. Compared to 2.4 GHz standards, 5 GHz standards have advantages such as greater scalability, better interference immunity, and higher speed. Those advantages allow for higher-bandwidth applications and more users. Table 2.3 shows the power regulations in the 5 GHz band. Here, EIRP refers to the peak output power delivered to the directional antenna in the strongest direction whereas maximum ratings (Max) indicates the peak output power delivered to the antenna [12].

Table 2.3 Power regulations in the 5 GHz band [12].

Frequency / Country	5.15 – 5.25 GHz	5.25 – 5.35 GHz	5.470 – 5.725 GHz	5.725 – 5.825 GHz
United State	50 mW (Max) 200 mW (EIRP)	250 mW (Max) 1 W (EIRP)	N/A	1 W (Max) 4 W or 200 W (EIRP)
Europe	200 mW (EIRP)		1 W (EIRP)	25 mW (EIRP)
Japan	200 mW (EIRP)	N/A	N/A	N/A

Different frequency bands may result in the different characteristics of the Wi-Fi signal including the fluctuation of the signal, the damping of the signal through walls. The use of 5 GHz for Wi-Fi fingerprinting has been studied by some research groups [31], [53]–[55]. They compared the standard deviation [31], [53] or the statistics [55] of received signal strength (RSS) values from 2.4 GHz and 5 GHz networks, and thus, infer the potential impact on the location accuracy (i.e., the error distance) of Wi-Fi fingerprinting. On the one hand, the coverage of such networks is considered. The coverage distance of a 5 GHz AP is smaller than the coverage distance of a 2.4 GHz AP when the radio transmission powers of the two devices are equal [12], [54]. Therefore, 5 GHz signals require more APs to cover the same area as for 2.4 GHz signals.

On the other hand, the signal stability is also investigated. For instance, Farshad et al. [31] calculate the mean and the standard deviation of RSS values from the different bands (e.g., 2.4 GHz and 5 GHz) of the same AP. The 5 GHz has lower mean RSS while the 2.4 GHz consistently has a higher standard deviation. The potential reasons for a more stable RSS of 5 GHz signals are that 5 GHz beacons are sent at higher bit-rate, and 5 GHz signals have low co-channel interference. The authors also study the impact of frequency band on Wi-Fi fingerprinting by using a smartphone to collect multiple 2.4 and 5 GHz samples for each location. As a result of their study, they conclude that including the 5 GHz band offers significant improvements in Wi-Fi fingerprinting accuracy because of lower signal variations compared to the 2.4 GHz. Similarly, Lui et al. [53] have investigated different chipsets operating on dual bands to test how different devices behave. The result shows that different devices at the same point reported the difference in mean signal strength. The differences reported signal strength in an indoor environment can be up to 30 dB. The big difference noted is also true for devices from the same manufacturer. The authors also compare the variation of 2.4 and 5 GHz signals. The standard deviation of 5 GHz signals is consistently lower than that of the 2.4 GHz signals. Therefore, the use of 5 GHz could potentially improve the accuracy of a Wi-Fi fingerprinting system due to its higher stability than for 2.4 GHz. The authors suggest performing calibrations across different devices to maintain reasonable accuracy.

However, some researchers present different results compared to the conclusions of the two groups above in their literature. For example, Talvitie et al. [55] studied the statistics of RSS values from the perspective of fingerprinting localization. They reported that the observed RSS values of 5 GHz networks are lower than the observed RSS values of 2.4 GHz. Suppose the high RSS values are crucial for the Wi-Fi fingerprinting, the location accuracy of using 2.4 GHz (with relatively high probability to receive high RSS values) should be better than using 5 GHz. Accordingly, their experiment results show that the localization performance with 5 GHz networks is worse than when using 2.4 GHz networks – more specifically, it results in worse accuracy and less floor wide detection probability. These results are contrary to those presented by the two groups introduced above. The reason for this is maybe because of the way Talvitie et al. utilized the measured data. In their experiment, to compare the positioning performances between the two frequency bands, the authors have limited the number of APs at each location by filtering the lowest RSS values of 2.4 GHz samples so that the number of samples measured at each location is the same for both of the frequencies. The authors do so to have comparable coverage areas for both frequency bands, and therefore the comparison of positioning result between the two frequencies become fairer than using the full database. However, the authors consider only the coverage of 2.4 and 5 GHz signals but do not consider the variation of signal strength. In reality, the coverage distance of 5 GHz APs is smaller than that of 2.4 GHz APs. Thus, at locations which are far from the APs, the measurement device can measure the signal of 2.4 GHz APs but cannot receive signal of 5 GHz APs. At those locations, the 5 GHz signals are weak, unstable and may fluctuate considerably. Therefore, in this study, the fingerprinting system has poorer accuracy when using the 5 GHz signals compares to using the 2.4 GHz signals.

Those studies considered only limited situations of radio propagation indoors such as the path loss in a hall or a large room without walls during the path. Indeed, in an indoor environment, buildings are often divided into rooms, very often with small dimensions (e.g., an office of 10-15 m^2). The radio propagation from an access point can cover several rooms. Thus, the radio propagation in the indoor environment is complex, not only because of the reflection and multipath caused by the walls around, but also the damping and scattering through the walls between them. In such scenarios, it is necessary to consider radio propagation in the areas with several small-dimension rooms.

The theoretical analysis and the experimental investigation of the signal coverage and stability in the literature help to understand the influence of radio frequency utilization on the performance of Wi-Fi fingerprinting. However, the results of fingerprinting localization systems utilizing 2.4 and 5 GHz are more persuasive, which is one of the goals of this thesis. Besides, the use of different frequency bands for Wi-Fi fingerprinting does not only influence the performance of localization but may also influence the cost of resources of the mobile devices, e.g., the power consumption.

2.4.3 Small-scale Fading

Another issue that influences the Wi-Fi signal strength is small-scale fading, which is caused by multipath propagation. Moving the measurement devices over a very short distance can experience small-scale fading and result in the severe fluctuation of the RSS values. As a consequence, the fluctuation of RSS values may also influence the performance of Wi-Fi fingerprinting. The influence of small-scale fading on the RSS values has been investigated in the literature [56], [57] but its influence on Wi-Fi fingerprinting has not been investigated yet. In [57], the authors investigate the susceptibility to the fading effect of the Bluetooth signal. The authors measured the Bluetooth RSS while gradually moving the measurement device (iPhone) up to 3 meters toward the AP. The result shows that the Bluetooth signal strengths fluctuate deeply in all channels when the measurement position is moved even just 10 cm. For the 3-meter distance, the signal strength varies up to 30 dB. The authors also try to mitigate the influence of small-scale fading on the performance of a Bluetooth fingerprinting system by comparing the accuracy of the system using the raw data and using the max, mean, and median value with a window length of 0.5 and 1 second. The result shows that the system achieves higher positioning accuracy when applying those three mitigation schemes. V. Moghtadaiee and A. G. Dempster [56] investigate the relationship between the geometric distance and the vector distance between a pair of reference points by analyzing data measured at a very short distance from each other at [0, 1, 5, 10, 20, 50, 100, 200] cm. The authors concluded that the RSS variation due to the small-scale fading effect is significant even when the geometric distance of the measurement positions is very small. However, when the Wi-Fi fingerprints are more than a set of Wi-Fi signal strength values, the influence of small-scale fading on the RSS-ranges, as well as on the performance of the Wi-Fi fingerprint systems are not well-investigated yet.

2.4.4 The influence of the presence of people on Wi-Fi signal strength

The Wi-Fi signals radiate over the air and are influenced by environment absorption, reflection, multipath, scattering, the presence of people, among other factors. The presence of people in the radiated environment causes the fluctuation of the Wi-Fi signal [58]. In [59], Kaemarungsi et al. reported that the distribution of received signal strength spread over a larger range when there was a human presence. The standard deviation increases from 0.7 dBm to 3 dBm. The obstruction of the body also attenuates the signal strength. The authors measured the Wi-Fi signal in different directions in which users stay in front of and behind the transmission line between the Wi-Fi AP and measurement device. Results show that there is a large difference in signal strength value in those cases.

To deal with the influence of the presence of people and other environmental factors on the signal strength value and the performance of a Wi-Fi fingerprinting system, researchers have proposed the use of sensors to detect changes in the environment and use different context-aware fingerprint databases to adapt to the environmental condition. Chen et al. [9] have proposed to use RFID and environmental sensors to detect changes in the environment. The humidity sensor is used to detect the humidity level, and the Bluetooth device is used to detect people presence. Based on the environmental condition, the system selects the fingerprint database which is best suited to the environment. Another approach is based on the Wi-Fi signal of the known APs measured by a device at a fixed location to aid in deciding the radio map that best represents the environmental condition [60].

2.4.5 Heterogeneous devices

Wi-Fi fingerprinting works in two phases: first, the measurement devices measure Wi-Fi signals at different places to generate the fingerprint of those places; then, in the positioning phase, the current Wi-Fi scan will be compared with the fingerprints generated in the first phase to figure out the current location. To maintain the good performance of the system, the environment as well as the devices used in those two phases should not be changed. However, in reality, the devices used to generate the fingerprints during the training phase may differ from the devices used during the positioning phase. This issue is known as the device heterogeneity problem. Such problems will influence the performance of a Wi-Fi fingerprinting system. The received signal strength P(d) at a specific distance d is given by the following equation [42], [61], [62]:

$$P(d) = 10\,log\left(\frac{G_{MN}\,G_{AP}P_{AP}\lambda}{(4\pi d_o)^2}\right) - 10\beta log(\frac{d}{d_0}) + X_1$$

G_{MN}: the antenna gain of the mobile device

G_{AP}: the antenna gain of the access point

P_{AP}: the transmitted power of the access point

λ : the carrier's wavelength

d_0 : the reference distance

β : the path loss exponent (depend on radio environments, ranging from 1.6 to 6 in typical environments)

$X_1 \approx N(0, \sigma^2)$: variation of the received power. It is a Gaussian distributed random variable with mean zero and variance σ^2

The above equation indicates that the signal strength depends on the antenna gain of the measurement device and access points. Different measurement devices may have different antenna gain values. Consequently, if different measurement devices are used in training and positioning phases, the signal strength value measured by them may also be different.

The influence of device heterogeneity on Wi-Fi signals has been studied. Several papers reported that different measurement devices with different Wi-Fi chipsets and antennas measure varying RSS values, even though the devices are placed at the same location. The authors in [53], [63] report that different Wi-Fi chipsets from different manufacturers perform differently, and therefore give different RSS values. The difference of the signal strength measured by different devices may be up to 30 dB [53]. Varying RSS values from different devices may influence the performance of Wi-Fi fingerprinting. Therefore, the calibration among different devices needs to be done to maintain good performance. To demonstrate the heterogeneity device problem, two mobile devices (Nexus 5 smartphone) were placed next to each other to measure the Wi-Fi RSS broadcasted from an AP. The RSS value is plotted in Fig. 2.5. The x-axis represents the number of measurement samples, and the y-axis represents the signal strength value. From this figure, we can see the difference of Wi-Fi signal strength measured by two devices at the same location. The different signal strength value of the two measurements is approximately 7 dB.

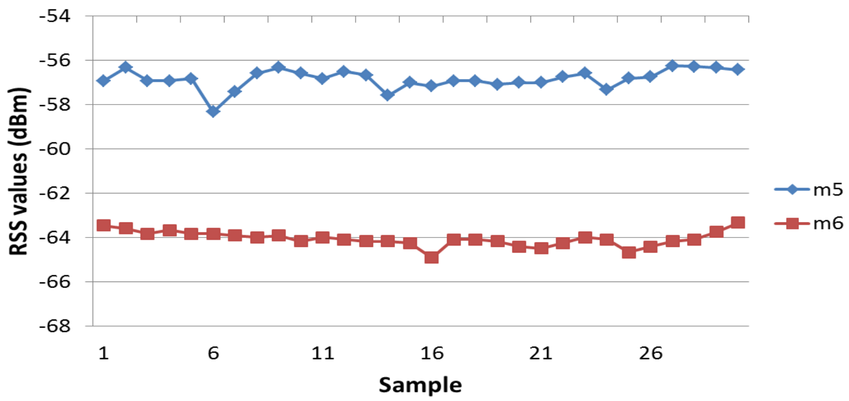


Figure 2.5 Wi-Fi signal measured by two devices at the same location.

2.5 Improving the performance of the Wi-Fi fingerprinting system

2.5.1 Using channel state information

The RSS is sensitive to temporal changes in the environmental factors [63], [64] and cannot help in understanding the variation of transmission channels. OFDM's multicarrier approach, used in 5 GHz Wi-Fi standard including 802.11a, 802.11n, 802.11ac standard, is an encoding scheme that uses many lower speed subcarriers to transmit data. Multiple subcarriers travel through different traveling paths and result in differing amplitudes and phases of each subcarrier. The channel quality of the transmission link between a transmitter and a receiver is described by channel state information (CSI) which reveals information at subcarrier level including magnitude and phase information. Therefore, CSI contains both the phase and amplitude of each subcarrier and can be captured to justify the multipath effect and better represent location information. Channel state information can help to understand the impact of channel fading, provide prominent features to describe unique location signatures and be utilized as the fingerprint for a Wi-Fi fingerprinting system. Currently, Wi-Fi interface on the smartphone does not support the collection of channel state information. Halperin et al. [65]

have developed a tool run on Linux operating system using the Intel 5300 network interface card (NIC) to extract the CSI.

Channel state information has been investigated by researchers evaluating Wi-Fi fingerprinting. In [66], Wang et al. utilized CSI information to establish a Wi-Fi fingerprinting system which collects CSI information from all subcarriers for all antennas. The system achieves higher accuracy compared to other methods despite using just one AP. The authors also investigate parameters which influence the performance of the system such as different antennas, number of packets, environment variation, and training grid size. Using all three antennas with 90 CSI provides higher accuracy compared to using only one of them. However, it takes more time to process data when using 90 CSI. The authors also change the distance from obstacles to the AP. The obstacle has less impact on the signal when the distance is increased. When the grid size is increased, their CSI properties have less similarity.

In [67], Chapre et al. have utilized the channel state information of subcarriers as fingerprint features instead of using RSS value in their Wi-Fi fingerprinting system. The authors aggregate the CSI from multiple antennas for each subcarrier to generate a complex CSI-MIMO signature and amplitude-based CSI-MIMO signature for the static and dynamic environment, respectively. The authors also examine the effect of various impact factors including the size of the CSI-MIMO signature, the number of training and testing samples, and number of APs on the performance of the system. They achieve the maximum accuracy of 0.98 and 0.31 m using the k-nearest neighbor algorithm in the static and dynamic environment.

2.5.2 Addressing device heterogeneity problem

Different methods have been introduced to deal with the problem of using heterogeneous devices in a Wi-Fi fingerprinting localization system. Manual calibration and linear transformation methods have been used to calibrate the positioning devices so that they can use the FP generated by another device instead of surveying and generating the FP by themselves. Several publications mention that the signal strength measured by different devices from the same AP follows a linear model [68], [69]. Haeberlen et al. use the linear transformation to calibrate the signals measured by different devices [70]. The authors assume that the difference of Wi-Fi RSS measured by different devices have a linear relationship and could be compensated by a linear transformation. The Wi-Fi signal measured by a reference device is used to generate the fingerprint, then an unknown device manually surveys the Wi-Fi signal in several locations which have already been measured by the reference device. Then, the parameters used for linear transformation between the signal strength measured by two devices at the same locations is calculated. Based on those parameters, the whole fingerprint database for the unknown device is generated. However, manual data collection and calibration are labor intensive work. This approach requires an amount of time and effort to measure the Wi-Fi signal in some reference places for each new device. Then, the relationship between each new device and the reference device must be mapped. The number of new mobile devices is increased considerably every year. Therefore, this approach seems to be unpractical in reality. Moreover, Park et al. state that a linear transformation alone cannot solve the heterogeneous device localization [71]. Instead, they suggest using kernel estimation with wide kernel width to mitigate the difference of signal strength across devices. Another study learns the linear transformation parameters by analyzing the Wi-Fi signal when the positioning device passes certain locations which are easy to recognize. In [72], the authors used the Pearson correlation ratio to compare the similarity of positioning estimates using different devices. First, the positioning device roughly estimates the location based on the current Wi-Fi signal and the fingerprints in the database. Then the Pearson correlation ratio between the current Wi-Fi scan and the estimated fingerprint is calculated. This ratio indicates the relationship between the training and the positioning devices and is used to calculate the parameter of the transformation function.

Besides, other approaches extract features which do not depend on the receiver hardware to use as location's FP such as using the differences of RSS values or the ratio of RSS values between pairs of APs, instead of using absolute RSS values, in order to eliminate the influence of device heterogeneity [42]–[44], [73]. Kjærgaard et al. [73] use the RSS ratio instead of using absolute RSS value in their study. The visible APs is ranked in order, and the ratio between two APs is calculated to form the fingerprint which consists of a vector of RSS ratios for each pair of AP. Another study uses RSS differences between pairs of APs as a fingerprint feature [42]. The difference between the signal strength of two APs is calculated and used as a feature to generate Wi-Fi fingerprint. Machaj et al. rank the list of APs in order based on their RSS value and use the rank order as a fingerprint feature [46]. Those features reflect the relative relationship between the pair of APs, so it can help to avoid the heterogeneous problem caused by using different devices.

2.5.3 Reducing energy consumption

Intensive use of GPS or Wi-Fi scanning consumes lots of energy and leads to the exhaustion of power in a short time [68], [69]. Several researches have investigated the power consumption of scanning Wi-Fi networks and proposed methods to reduce power consumption. Two approaches have been applied to save power: reduce the frequency of Wi-Fi scanning and reduce the number of APs scanned.

- Reduce the frequency of Wi-Fi scanning: for localization purposes, the frequency of Wi-Fi scanning can be reduced when the user is not in motion state. Accelerometer sensor of the mobile device can be utilized to detect the motion state. If the state is static, the device stops the Wi-Fi scanning until the moving state is recognized [68], [74]–[76]. Xu et al. [77] have proposed a power-save strategy for which learning and positioning occur only when the system detects the mobile devices are in a stationary state based on the built-in accelerometer sensor.

- Reduce the number of AP scanned: Brouwers et al. [78] have proposed an energy-efficient Wi-Fi scanning algorithm to reduce energy consumption as only a subset of channels is scanned. This method focuses on the useful channels of APs for the localization. The authors observe that some channels are more popular than others. Moreover, a limited number of channels are enough to provide high accuracy for the system. Therefore, instead of scanning Wi-Fi channel as in the order channel list, the device should scan popular channels first and continue the scanning process until it finds a sufficient number of APs for localization purpose. By doing that, the scanning time and energy consumption for the scanning process can be reduced. The authors have proved that their method is an efficient approach to save energy. However, this method requires modification of the device's operating system so that it can scan the selective channels. Also, it requires to know the popular list of APs which can be different in different environments.

Most of the approaches mentioned above are developed in the scenario of 2.4 GHz networks. As one of the performance measures for the system reliability in practice, the power consumption demanded when running the fingerprinting localization systems in the scenarios of 2.4 GHz and/or 5 GHz networks needs to be well-considered.

2.6 Evaluation Metrics

Different evaluation metrics have been used to assess the performance of an indoor localization system.

2.6.1 Accuracy and Precision

The accuracy and precision of the metrics are used popularly in evaluating the performance of a localization system. Those two metrics are defined interchangeably in different publications. Several systems report the accuracy in meters as how close the estimated location is near to the ground truth. On the other hand, other systems report the accuracy as the percentage value of the number of times that a system classifies the location correctly. A higher accuracy indicates a better system. Besides accuracy, precision is another common metric used to evaluate the performance of a localization system. The metric precision is usually presented as the cumulative density function (CDF) which states how consistent the system works. Precision indicates the distribution of the distance error between the estimated location and the ground truth. In [79], Wagner et al. defined accuracy as the percentage of correct classifications between the estimated location and the ground truth. Precision is defined as the difference in the value of the estimation compared to the ground truth in meter. Kjærgaard in [73] and Jiang et al. [5] utilized a similar definition of accuracy as Wagner. In [19], Swangmuang et al. defined accuracy as the error distance between the estimated and the actual locations; whereas precision is defined as the percentage of successful location estimates with a given accuracy. Fuchs et al. defined precision as the average absolute positioning error [80]. In [81], the metric accuracy is used as the percentage of correct predictions within some error distance. When two systems have similar accuracy, the system with higher precision is better. For instance, when one system has a precision of 60% within 2m and 90% within 3m while another system has a precision of 80% within 1.8m and 93% within 3m, the latter system has a higher precision than the former system.

2.6.2 In meter or room-level accuracy

Many studies presented the performance of their system in term of meter accuracy. Meter accuracy indicates how the inferred location nears the ground truth. The accuracy of those systems is reported is 2-3 meters, some others even reported the accuracy of less than 1 meter. In [82], the authors mentioned that the accuracy increases from 7.2 meters to 4.8 meters when the number of APs increases from 1 to 2. They achieved 2.56 meters of accuracy when 10 APs are used. Other publications described the accuracy of a Wi-Fi fingerprinting system as room-level accuracy. The systems aim to distinguish the room or the floor level. Several publications concluded that the location of a user within a room is sufficient for most context-aware applications [83], [84]. For indoor localization, it is useful if the system can differentiate user in the room they are in to suit their needs. The building is designed with different rooms having varying functionalities such as living room, bedroom, kitchen room, etc. People occupy different rooms which have different purposes, hence it is necessary and enough to distinguish between such rooms.

2.6.3 Complexity

The complexity of the localization algorithm is also an important factor which needs to be considered when designing a localization system. When mentioning the complexity of a system, the complexity of an algorithm is of more concern. For instance, people may design an algorithm to run on a mobile device or on a centralized computer. The hardware capability of a mobile device is limited and cannot be as powerful as that of a personal computer. The power of the mobile device is also limited. Therefore, if an algorithm runs on a mobile phone, the complexity of the algorithm should be low. Otherwise, it takes much more time to perform a positioning task. The computing time is important for a localization system. After sending the location request, a user may move to another place. Thus, if the system requires a long time to answer the location request, the answer will not be correct anymore because the user has left that location.

Some positioning systems design the algorithms to run completely on a mobile phone. Other systems combine both the mobile phone and the computer to calculate the location. The mobile device measures data and sends the data to the centralized computer. Then the computer performs the positioning function and returns the result to the mobile. If the algorithm runs on a computer, it will result in shorter the processing time, but the mobile and the computer need to communicate and transfer data regularly. For example, the Active Bat system [85] sends all information to the computer to do the data analysis for tracking the location of the transmitters.

2.6.4 Recognition rate

Besides the requirement of location accuracy, users also want to get such accurate localization most of the time. The recognition rate indicates how fast or how often the localization system can position a mobile device / a user. When a Wi-Fi fingerprinting system compares the current Wi-Fi scan with the fingerprints to get the location, it may figure out the location or not. The recognition rate compares the times the system answers the location request. In this thesis, besides accuracy, I also use the recognition rate to compare the performance of our fingerprinting system in different scenarios.

2.6.5 Power consumption

Mobile devices are often powered by a small battery. Therefore, the energy source of a mobile phone is limited and does not last for a long time. Although the battery capacity of the mobile phone has drastically increased in recent years, the functions which a mobile phone does also increase, hence saving power is still an important objective. The Wi-Fi fingerprinting approach necessitates scanning of a Wi-Fi signal frequently in order to perform the localization function. In this way, locating a users' position is bound to consume a lot of energy. Therefore, it is necessary to save energy while providing positioning services; otherwise, the positioning function drains the energy source very fast.

Several researches have investigated the power consumption of Wi-Fi scanning tasks and have proposed methods to save energy. In [78], Brouwers et al. investigated the power consumption of different mobile phones related to scanning Wi-Fi signal and suggested a partial scanning approach saving energy. This approach first learns the most popular Wi-Fi channels used and the number of APs enough for a Wi-Fi fingerprinting system. Based on this information, the partial scanning approach scans the most popular channel first instead of scanning channels as in channel list orders. It stops scanning when it gets a Wi-Fi signal from enough number of APs. Compared to scanning all Wi-Fi channels, this approach helps to save energy as it scans fewer Wi-Fi channels. Other researches leverage the built-in mobile sensors to detect whether the mobile is in stationary or motion state to start or stop scanning Wi-Fi signal. Xu et al. [77] proposed a power-save strategy for which learning and positioning occur only when the system detects the mobile devices are in a stationary state by basing on the built-in accelerometer sensor. Shafer et al. [76] developed a strategy to perform full localization only when the mobile device has detected that a user has finished moving to another location.

Faragher et al. [86] introduce SwiftScan, a Wi-Fi fingerprinting scheme that can reduce the number of Wi-Fi channels scanned to reduce energy consumption. This method performs the site survey and stores the detail of Cell-ID of a cellular network, Wi-Fi channel and MAC address in a lookup table. Then, based on the currently observed cell-ID, the positioning device scans the most popular Wi-Fi channel first to look for known MAC addresses. The presence of any known APs confirms the user is in a known location. If the device does not see any known APs, the scanning process is continued with other channels in the channel list in the lookup table until it meets the minimum required number of APs. This helps to minimize the number of scanned channels.

2.7 Analysis methods and tools

2.7.1 Analysis of Variance test (ANOVA) and t-test

ANOVA is a popular statistical hypothesis test to evaluate whether the mean values of multiple groups are different or not. The null hypothesis for ANOVA is that all group means are equal. In general, ANOVA compares variances among groups and variances within the group. The outcome of the ANOVA is the p-value. Based on the p-value and the significance level, people can determine whether the null hypothesis is rejected or accepted. If the p-value is smaller than or equal to the significance level, we can reject the null hypothesis. In other words, we can conclude that not all the group means are equal.

The result of the ANOVA test signifies that the group means are not all equal, but it does not specify which means are different. To have a more specific result of which pair of data is similar or different, the t-test between a pair of data can be used. The t-test is a common statistical hypothesis test used to compare the means of one or two data sets [87]–[90]. The t-test investigates two mutually exclusive hypotheses: the null hypothesis (H_0) and the alternative hypothesis (H_1). The t-test assesses the sample data to decide which hypothesis the data supports. The null hypothesis states that there is no difference between the group means; the alternative hypothesis indicates that there is a difference between the group means.

- H_0: $\mu_1 = \mu_2$ ("the two-group means are equal")
- H_1: $\mu_1 \neq \mu_2$ ("the two-group means are not equal")

T-test uses a specific procedure to calculate the sample data and produces the result as a standardized t-value. The calculations behind t-value consider both the variability of the data between groups and within the group. The variance of data within group measures or describes the dispersion of samples around its group mean. Therefore, t-value or t-test describes the difference of two datasets better than the normal comparison of mean values. If the different mean of two groups is small relative to the variance within each group, we have a low t-value. If the different mean of two groups is large relative to the variance within each group, we have a high t-value. To reject the null hypothesis, we need a high t-value.

$$t - value \ = \frac{variance\ between\ groups}{variance\ with\ group} = \frac{mean\ difference}{standard\ error}$$

$$standard\ error = \frac{standard\ deviation}{\sqrt{number\ of\ sample\ data}}$$

The t-value indicates the difference or similarity between the two groups. A t-value of 0 means there is completely no difference between two groups or the sample results exactly equal the null hypothesis. If the absolute value of the t-value increases, the difference between the sample data increases. However, a single t-value is difficult to interpret if it is high enough to reject the null hypothesis. To interpret it, people need to place t-value into a larger context of t-distribution to calculate the probability. Every t-test produces a single t-value. If we performed the same t-test for multiple random samples of the same size from the same population, we would obtain many t-values. Then, we could plot a distribution of all those t-values which is known as a t-distribution. Fortunately, the properties of t-distributions are well understood in statistics. Therefore, we do not need to collect many samples to draw the t-distribution. The t-distribution has the probability density function given by the following equation [90]:

$$f(t) = \frac{1}{\sqrt{v}\, B\left(\frac{1}{2}, \frac{v}{2}\right)} (1 + \frac{t^2}{v})^{-\frac{v+1}{2}}$$

B: beta function

v: degree of freedom

The beta function B is defined as:

$$B(x, y) = \int_0^1 t^{x-1}\,(1-t)^{y-1}\,dt$$

Until now, we know t-value and t-distribution. To decide whether t-value is unusual enough to reject the null hypothesis, we need to place a t-value in the context of the correct t-distribution to calculate the probabilities associated with that t-value. This probability is called the p-value. The outcome of the t-test is the p-value which indicates how high the data contradicts the null hypothesis. Lower p-value provides stronger evidence against the null hypothesis. The significance level and the p-value are used in conjunction to decide which hypothesis the data support. The result of the ANOVA test and the t-test is considered significant if the p-value is less than the significant level α. If the p-value is smaller than or equal to the significance level, we may reject the null hypothesis.

To assess the ANOVA and t-test result, we need to define the significant level α we want to accept. The significant level indicates how strong evidence we need to reject the null hypothesis for the entire population. Lower significance levels indicate that we require stronger evidence before we reject the null hypothesis. The significance level is selected before running data analysis, and typically set to 0.05 which indicates a 5% probability of mistakenly rejecting the null hypothesis whereas the null hypothesis is true.

Fig. 2.6 demonstrates the t-distribution with the degree of freedom equals 20. The peak of the t-distribution is right at zero, which indicates that there is the highest probability to obtain a sample value close to the null hypothesis. For example, we perform a two-tailed t-test with t-value equal 2. A two-tailed t-test evaluates the difference between two groups is statistically significant in either the positive or negative direction. Then, the p-value is the cumulative distribution of the area of the curve that has t-values greater than 2 and t-values less than -2. In this example, the total probability is 0.05926. This p-value is higher than the significant level α 0.05, so we cannot reject the null hypothesis. In other words, the means of the two groups are not different.

In this thesis, ANOVA and t-test are run to see whether the average RSS values of data sets in small-scale fading and device heterogeneity experiment are different or not. In practice, we can use available functions in Microsoft Excel or MATLAB as well as other statistical applications to quickly perform the ANOVA test and t-test.

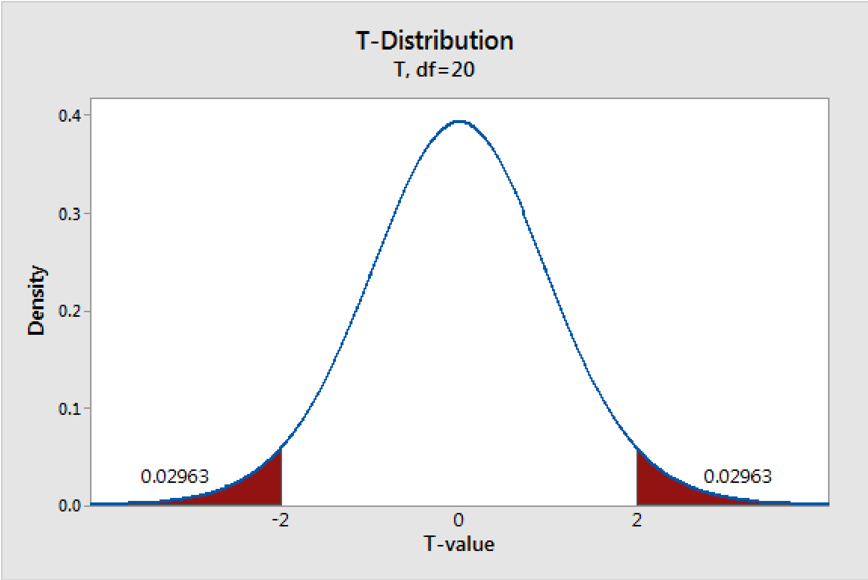


Figure 2.6 An example of t-value and t-distribution [90].

2.7.2 Histogram

The histogram provides specific information on the distribution of the data. It demonstrates how many sample data falls into each statistical value or range of value. The histogram can also be used to show the distribution of data as relative frequency – the proportion of data in each value or range of value. In this case, the total distribution equals 100%. The histogram provides an intuitive understanding of the distribution of the whole data. In analyzing Wi-Fi RSS value, the histogram is also used to investigate whether the data follows the Gaussian distribution, right skew or left skew [63]. In this thesis, we plotted the histogram of the Wi-Fi signal to compare the distribution of 2.4 and 5 GHz signals, the distribution of signal when there are people and when no people are around the measurement area.

2.7.3 Box-and-whisker plot

The box-and-whisker plot describes the median, min, max value, the lower quartile, and the upper quartile of a set of data. Quartiles divide the data set into four equal parts; each part shows 25% of the data. The central box covers the first to the third quartile to show the middle 50% of data points. A line inside the box indicates the sample median. If the box appears as a single line, it indicates most of the data contains the same single value for that group. The whiskers extend from the quartiles to the smallest and largest data values in each sample.

Unusual points far from the box are indicated by a plus sign. In this thesis, Box-and-whisker plot is utilized to describe the distribution of Wi-Fi RSS value of various experiment [89], [91].

2.7.4 Weka tool

Weka is a popular machine learning tool which is used to analyze experimental data and run machine learning algorithms. Besides, Weka also provides free library for development purpose. This thesis utilized Weka to generate five-fold cross-validation data to validate our analysis result.

2.7.5 MATLAB tool

MATLAB is a powerful tool that performs various functions. In this thesis, Matlab is used to run ANOVA and t-test to compare the mean value of two or more datasets.

2.7.6 WHERE

The density-based clustering algorithm (DBSCAN) [15] is widely used to group a set of related data into clusters, which have a higher density than its neighbors. The higher density of a given cluster means the points inside the cluster are closer than points outside the cluster. The DBSCAN algorithm requires to define two parameters: the minimum number of points (MinPts) and the radius of the neighborhood (Eps). An individual point will be added to the group if the distance between them is smaller than or equal to Eps. A group of point of data will be recognized as a cluster if the number of points in the group is higher than MinPts. The key idea of the DBSCAN algorithm is that for each point p of a cluster, the neighborhood within a given radius Eps has to contain at least a minimum number of points MinPts, indicated as the neighborhood of p. If any two clusters contain the same point(s), the two clusters are merged into one cluster. For a point p, if it does not belong to any cluster, that point is regarded as noise.

WHERE applies DBSCAN on each AP to discover high-density clusters of RSS values which indicated as RSS-range. Then, clusters of RSS values from several APs measured at the same time, the same place are used as the fingerprint of that place. In this approach, each RSS value measured from an AP is considered as a point. The number of points appears in each RSS value is count and constitute a group of points. A cluster is a collection of consecutive points which has high density. In other words, a cluster is a set of RSS values and is indicated as an RSS range. Fig. 2.7 is used as an example to illustrate how WHERE applies DBSCAN to discover Wi-Fi RSS clusters. In this example, the neighborhood range Eps is set to 2, and the minimum number of point in a cluster is set to 120. First, the Wi-Fi data measured from each AP is sorted according to the RSS value. Then, the algorithm continuously scans through all the RSS value from the smallest to the largest value to detect clusters. In the range of [-60, -58] dBm, the number of points is eight which is smaller than MinPts, so this range is not considered as a cluster. The algorithm continues with the other RSS values. In the range of [-55, -53] dBm, the number of points is 157 which is higher than MinPts, so this range is considered as a cluster. The algorithm continues with the next value, and if the next range satisfies the cluster criteria, it is added to the current cluster. Finally, the algorithm expands the cluster to the range of [-55, -51] dBm. The number of points in the range [-52, -50] dBm is 84 which is smaller than MinPts, so the RSS value -50 is not included in the cluster. The same process was repeated with the other RSS values. In this example, the algorithm discovers one cluster, and the range of this cluster is later used to constitute the fingerprints of location.

The fingerprints of all locations are stored in the fingerprint database. Then, in the positioning phase, WHERE compares each RSS value of the current Wi-Fi data measurement with RSS-range of the fingerprints in the database to figure out the most suitable fingerprint which indicates the current location of the user. The most suitable fingerprint is the one which has the highest percentage of matching APs among all sensed APs of the Wi-Fi scans is return as a positioning response. More details about learning places using density-based clustering are introduced in the previously published papers [3], [4], [27], [77], [92], [93].

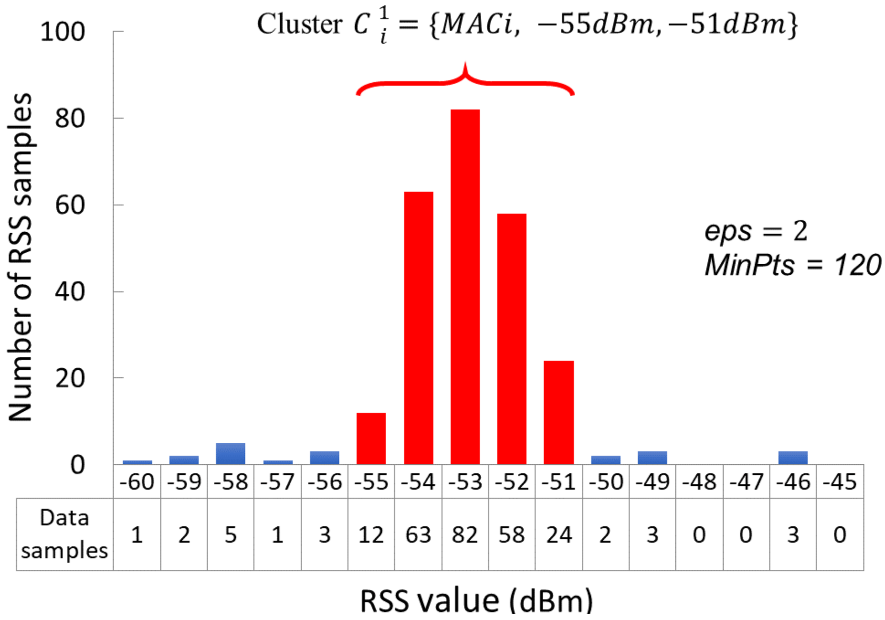


Figure 2.7 WHERE applies Density-based cluster algorithm to discover high density clusters of RSS values

2.8 Summary

In this chapter, I discussed issues related to positioning applications and techniques, its principle and challenges. Because of the inefficiencies of the GPS system in an indoor environment, diversity of approaches and techniques were studied to implement an indoor positioning system. Wi-Fi fingerprinting has become a promising approach for indoor positioning because of the availability of infrastructure and ease of the deployment. Then, the knowledge related to a Wi-Fi fingerprinting system was elaborated including different technology and techniques used in a fingerprinting system, site survey approaches to collect Wi-Fi data to generate Wi-Fi fingerprint of reference locations, features to use as fingerprints, and positioning algorithms. I examined various aspects which may influence a Wi-fi signal, the performance of a fingerprinting system, the difficulty in implementing a Wi-Fi fingerprint system, and the methods, techniques to improve the performance of the system. I also identified the evaluation metrics used to asset the performance of a fingerprinting system, analysis methods and tools used in this Ph.D. thesis.

3 Fluctuation of Wi-Fi Signals in an Office Environment

Although the theoretical model of radio wave propagation may describe and help to predict the radio signal strength at different distances, it does not guarantee that the model is validated in the real world. This is especially true for an indoor environment where the radio wave is influenced by various factors such as multipath, small-scale fading, scattering, reflection, the presence of people, the damping of the signal through a wall, and so on [94]. In this chapter, I examine the fluctuation of the 2.4 and 5 GHz Wi-Fi signal in a typical office environment. I carry out case studies in different areas of an office building including a hall, a corridor, office areas with four adjacent rooms. These areas are the common parts of buildings indoors. Smartphones (i.e., Nexus 5) are placed in the collection location to collect the Wi-Fi data with a frequency of 0.2 Hz. The collection duration at each location is half an hour. Identical tests were repeated for both cases: measure the Wi-Fi signal with the presence of people and without the presence of people. Details of the measurement setup and results are explained below.

3.1 Introduction

When radiating through the air, a radio signal's strength depends on several factors. The radio wave propagation model describes that the signal strength of a wireless signal at a specific distance depends on the frequency of the signal. Higher frequency results in higher free space path loss, which indicates how much signal loss is in the air [95]. Besides, different transmitted power also results in differences in signal strength. Although the 5 GHz signal has higher free space path loss than the 2.4 GHz signal, the 5 GHz and 2.4 GHz signal are allowed to transmit different maximum power. Water is another factor that may reduce signal quality because water can absorb the radio signal. The human body consists of water and can absorb the radio signal when they obstruct the signal path [58]. The level of absorption is also dependent on the frequency of the signal. The 5 GHz Wi-Fi signal is absorbed less when passing through water than the 2.4 GHz signal [96]. Therefore, a 5 GHz signal is less likely to be affected by wet objects and can penetrate walls, or human bodies better than the 2.4 GHz band. When facing obstacles, the radio signal must penetrate or scatter its signal around obstacles. The penetration ability of a signal increases with its frequency. Comparing to the 2.4 GHz band, the 5 GHz signal has the ability to penetrate a wall, ceiling with less attenuation. Higher radio frequencies also have better scatter characteristic [96], [97]. In a Non-Line-of-Sight (NLOS) indoor environment, the 5 GHz

Wi-Fi signal can scatter inside the building better than the 2.4 GHz signal. However, the scattered signal could result in signal reflections received from multiple, indirect paths and lead to a reduction in quality of the received signal. Therefore, the radio propagation model will not always describe the signal strength of a radio wave.

Besides the 2.4 GHz signal, the use of 5 GHz signal for wireless LANs is becoming more and more popular. It is worth investigating the characteristic of those two different Wi-Fi frequency bands in an indoor environment. The information would provide a better understanding of the 2.4 and 5 GHz signal which is utilized for Wi-Fi fingerprinting. In this section, the fluctuation of a Wi-Fi signal in corridors, halls, and office rooms was studied. The fluctuation of the signal is compared with the consideration of using different frequency bands, with and without the presence of people during the measurement. I also examine the degradation of signal strength transmitted through a wall.

3.2 Fluctuation of Wi-Fi signal in a corridor

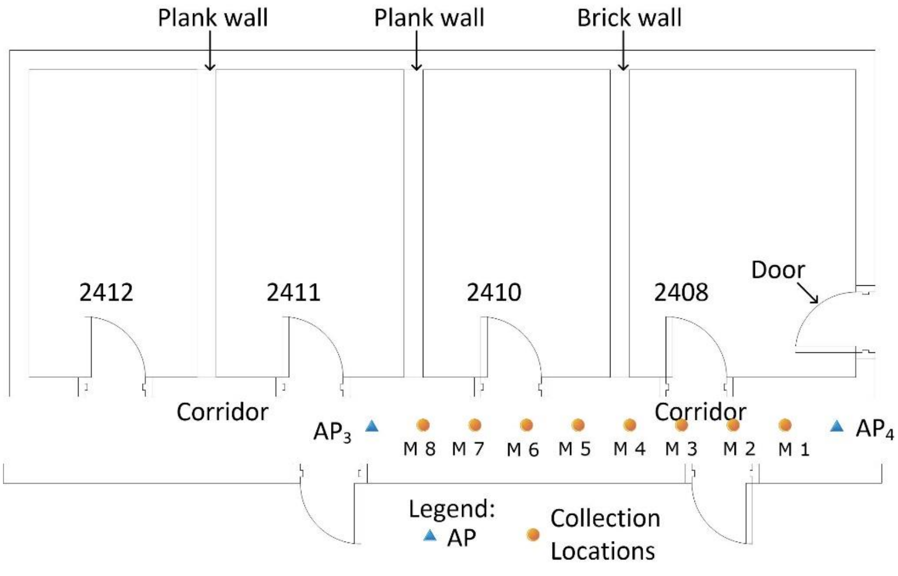


Figure 3.1 Layout of measuring signal in a corridor.

I carry out the experiment to investigate the fluctuation of Wi-Fi signal strength in a corridor. The corridor is a common part of an office building. Smartphones were placed to measure the Wi-Fi signal along the corridor with different distances to the APs; each location is one meter apart in space. The data is measured for 30 minutes. In this study, the APs were set up to broadcast both 2.4 GHz and 5 GHz frequency bands simultaneously. Thus, measurement devices receive both 2.4 and 5 GHz signals from the same AP. This helps to compare the fluctuation of 2.4 and 5 GHz signals. The layout of the experiment is shown in Fig. 3.1. There are walls between two sides of the corridor, so more reflection and multipath occur in such a scenario. First, I measure the signal when there are no people around the measurement area. Then, the same measurement procedure was repeated with the presence of people in the measurement area. The purpose of this is to compare the fluctuation of Wi-Fi signal with and without the presence of people during the measurement.

For the analysis, the measured signal was averaged and calculated the standard deviation. The purpose was to smooth the data and compare the fluctuation of the signal. The histogram of the measured data is also plotted to demonstrate the distribution of the signal. Table 3.1 and Fig. 3.2 show the mean and standard deviation (SD) of the measured data when there are no people around. According to the radio propagation path loss model, the RF signal strength decreases its value when the transmitted distance increases. Higher frequency signal has a higher loss. However, in our measurement, the analysis result shows that in the same measurement position, the RSS of 5 GHz signal may be weaker or stronger than the RSS of the 2.4 GHz signal. Moreover, the signal strength does not always decrease when the distance increases. The reason for this is probably because the multipath or reflection of the transmitted environment has caused the strange fluctuation of 2.4 and 5 GHz signal. We may observe in many cases that at the same measurement locations, the 5 GHz signal strength is stronger than the 2.4 GHz signal. Another interesting thing is the standard deviation of the 5 GHz signal is smaller than that of the 2.4 GHz signal in most cases.

Table 3.1 The mean and standard deviation (SD) of the 2.4 and 5 GHz signal measured in a corridor without the presence of people (Unit: dBm).

Locations	**AP3_2.4GHz**		**AP3_5GHz**		**AP4_2.4GHz**		**AP4_5GHz**	
	mean	***SD***	***mean***	***SD***	***mean***	***SD***	***mean***	***SD***
M1	-65.52	1.24	-61.01	0.31	-48.74	2.60	-43.99	0.38
M2	-67.01	0.87	-66.67	0.62	-52.28	0.71	-55.21	0.41
M3	-56.05	1.19	-61.92	0.36	-65.93	0.96	-56.85	0.36
M4	-55.82	0.58	-58.92	0.27	-59.40	1.82	-52.01	0.40
M5	-59.53	0.86	-58.69	0.46	-57.60	0.52	-50.94	0.23
M6	-63.12	0.76	-60.92	1.76	-66.60	1.98	-64.77	0.61
M7	-50.48	0.52	-55.44	0.50	-71.24	2.40	-62.21	0.50
M8	-52.67	0.60	-51.86	0.34	-77.38	1.30	-62.99	0.11

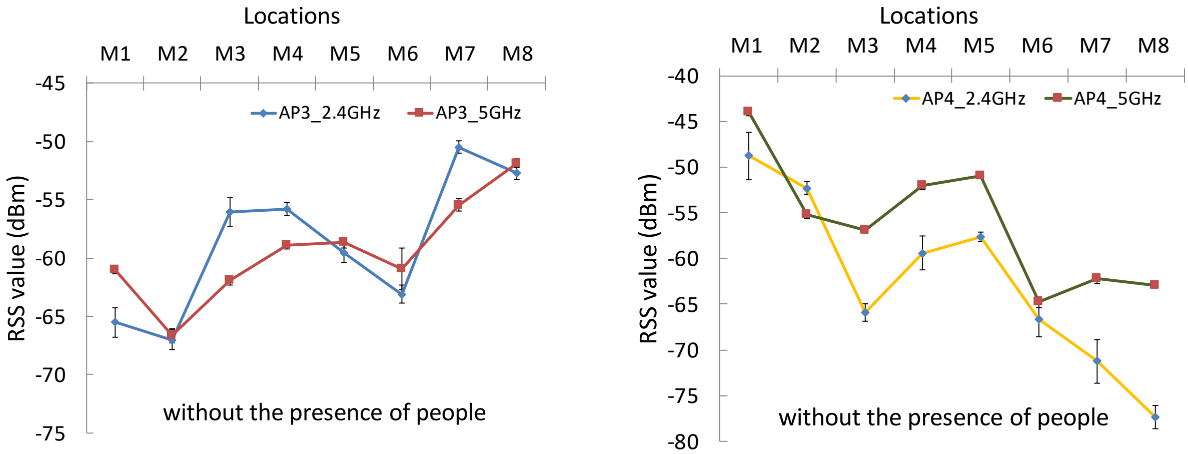


Figure 3.2 The signals measured in a corridor without the presence of people.

The same analysis procedure was applied to the data measured with the presence of people. Table 3.2 and Fig. 3.3 show the average and standard deviation of the signal. In this case, the signal strength and SD of a signal measured at different locations show the similar pattern as in the case there are no people around. It means the signal strength of the 2.4 GHz signal does not always higher than that of 5 GHz signal; the SD of 2.4 GHz signal is larger than the SD of 5 GHz signal. The difference between the two measurements is the fluctuation of the signal. When there are people around, the average signal at each location is similar to the signal strength in the case there are no people around; however, the fluctuation of the signal is much higher than the case there are no people around. The SD of signal when there are people around is higher than the SD of signal when there are no people around. The histogram of a signal measured with and without the presence of people shown in Fig. 3.4 demonstrates the influence of the presence of people on the signal strength. When there are no people around, the distribution of signal concentrates its value over a smaller range whereas with the presence of people, the signal distributes its value over a larger range. This is true for both 2.4 and 5 GHz signal.

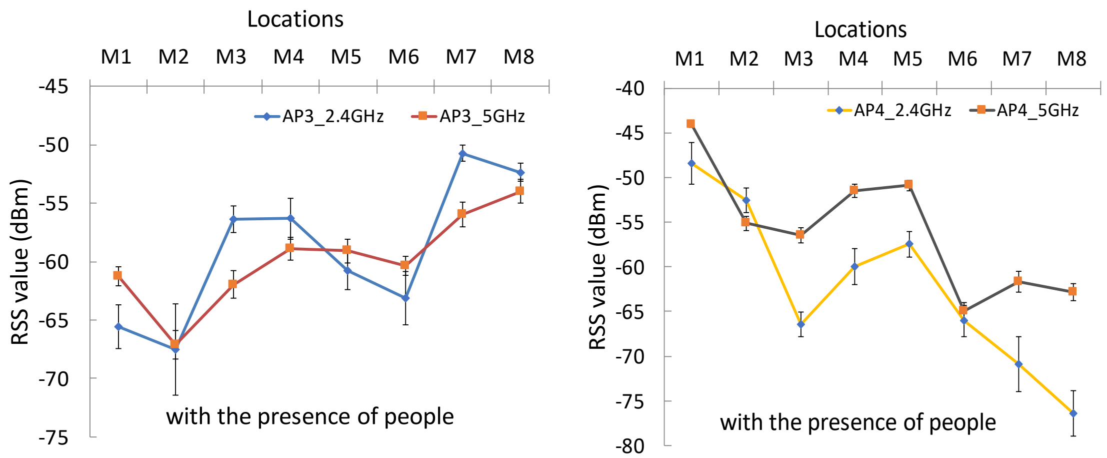


Figure 3.3 The signals measured in a corridor with the presence of people.

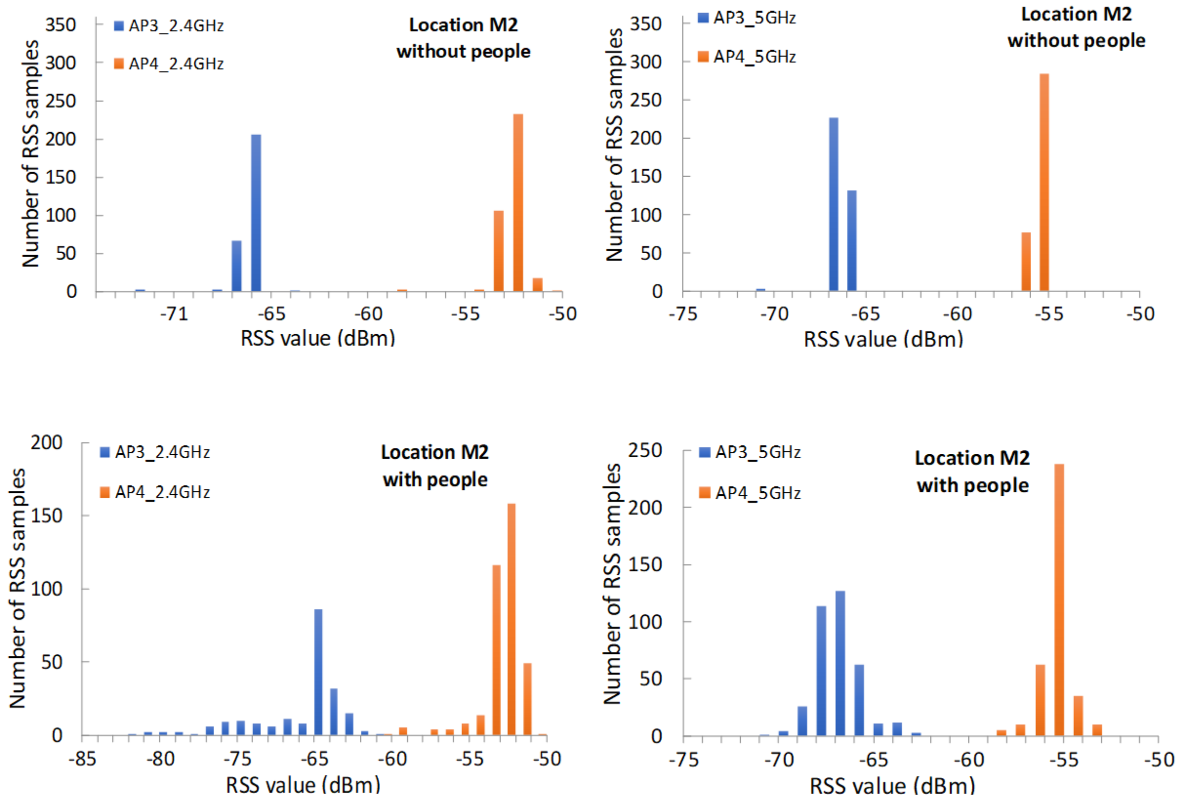


Figure 3.4 The histogram of the 2.4 and 5 GHz signals measured in a corridor.

3.3 Fluctuation of Wi-Fi signal in an office

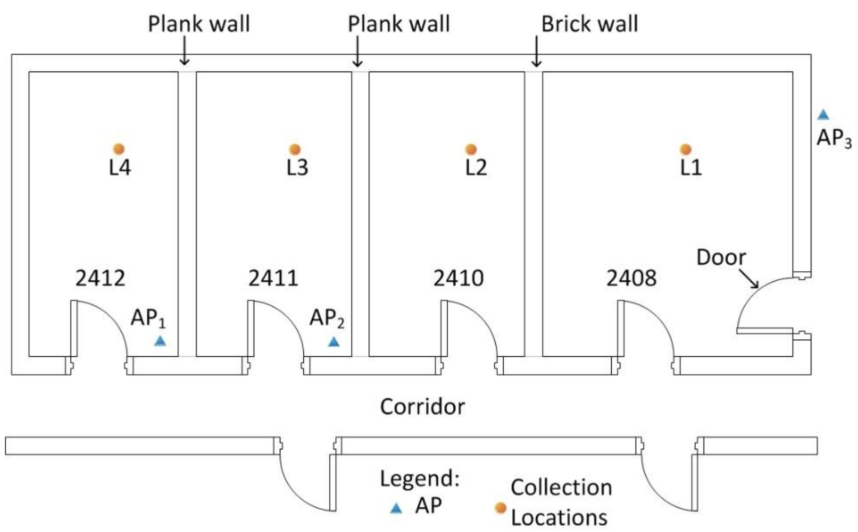


Figure 3.5 Layout of measuring signal in a corridor in office rooms experiment.

This experiment aims to examine the fluctuation of Wi-Fi signal in an office area consisting of different rooms separated by a wall and office equipment as obstacles. The layout of the experiment is shown in Fig. 3.5. Room 2408 has an area of 38.3 m^2, and each of the other three rooms has an area of 18.4 m^2. APs broadcast both 2.4 and 5 GHz signal simultaneously. First, the mobile phones collect Wi-Fi signal for 30 minutes when there is no presence of people. The measured data were averaged and calculated the standard deviation. Second, to compare the signal under the influence of people present, I collected signal when there is the presence of people. The similar analysis was done for this measurement.

The histogram of 2.4 and 5 GHz signal are plotted in Fig. 3.6. Two upper images show the fluctuation of the signal in case there were no people around; two lower images show the variation of the signal in case there were people around. We can see that in both cases, the fluctuation of the 5 GHz signal is smaller than that of the 2.4 GHz signal. Comparing the two images on the left side or on the right side, the signal fluctuated more when there were people around. In other words, the signal distributes its value over a larger range when there are people. This observation holds true for both 2.4 and 5 GHz band. Table 3.1 and 3.2 provides detail information about the mean and standard deviation of 2.4 and 5 GHz signal when there were people and no people around during measurement, respectively.

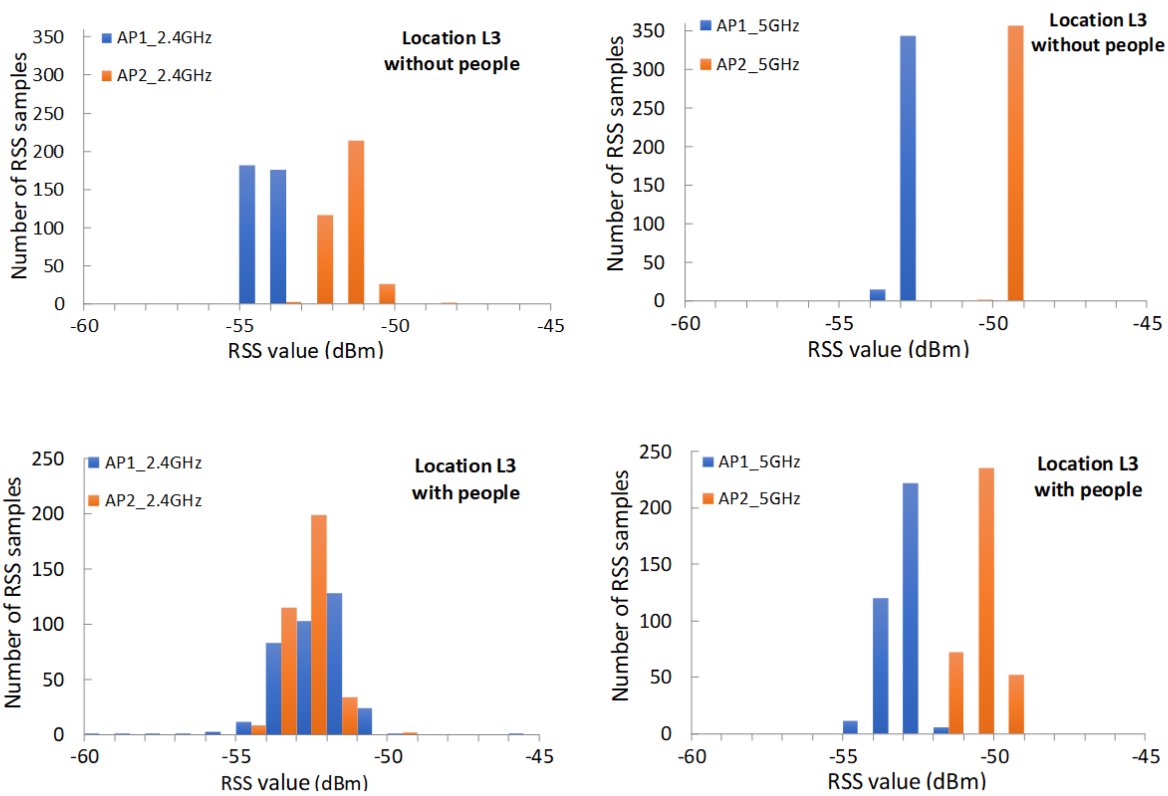


Figure 3.6 The histogram of the 2.4 and 5 GHz signals measured in office rooms.

Table 3.2 The mean and standard deviation (SD) of signals in different locations without people (Unit: dBm).

Access Point	L1		L2		L3		L4	
	mean	***SD***	***mean***	***SD***	***mean***	***SD***	***mean***	***SD***
AP1_2.4GHz	-69.15	1.11	-64.70	0.89	-54.53	0.61	-52.89	0.41
AP1_5GHz	-75.85	0.51	-65.48	0.50	-53.04	0.20	-52.04	0.30
AP2_2.4GHz	-64.65	1.01	-51.19	0.50	-51.25	0.62	-55.82	1.87
AP2_5GHz	-72.09	0.29	-58.10	0.30	-49.01	0.07	-55.42	0.49

Table 3.3 The mean and standard deviation (SD) of signals in different locations with people (Unit: dBm).

Access Point	L1		L2		L3		L4	
	mean	***SD***	***mean***	***SD***	***mean***	***SD***	***mean***	***SD***
AP1_2.4GHz	-70.23	1.73	-64.38	2.87	-52.87	1.25	-52.30	2.98
AP1_5GHz	-75.89	0.60	-66.30	1.00	-53.38	0.58	-51.39	1.15
AP2_2.4GHz	-64.69	1.49	-50.84	1.97	-52.36	2.16	-58.10	3.07
AP2_5GHz	-72.48	0.54	-58.02	0.30	-50.06	0.59	-57.32	0.47

3.4 The degradation of the Wi-Fi signal transmits through a wall

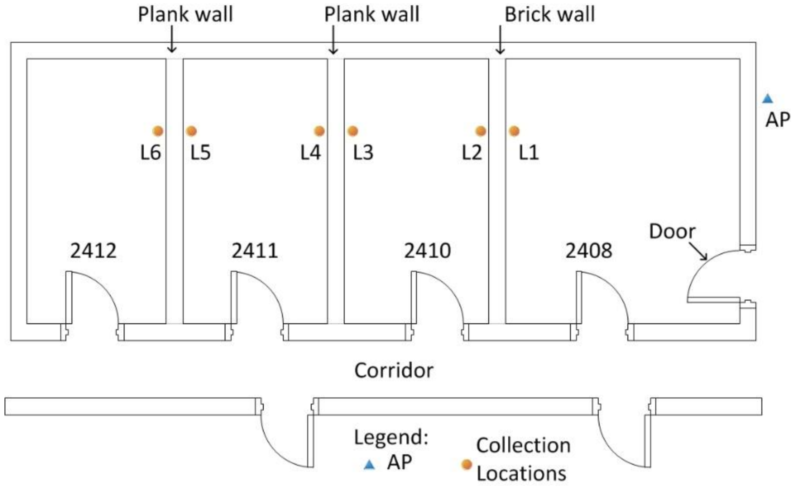


Figure 3.7 Layout of measuring signal in damping through wall experiment.

Radio frequency signal attenuates its signal strength depending on the distance between the transmitter and the receiver, the frequency of the signal and the obstructing materials type [98]. In our previous study [92], we observed that the 2.4 GHz Wi-Fi signal decreases its strength when it transmits through a wall. This experiment aims to investigate the difference of degradation strength in 2.4 and 5 GHz signal through different types of walls.

In this experiment, one AP was used to transmit the Wi-Fi signal in both 2.4 GHz and 5 GHz bands. A pair of mobile phones was placed in two sides of the wall to measure the signal for two hours (Fig. 3.7). Two types of wall appear in our experiment: brick wall and plank wall. A brick wall is the concrete wall of the building, and a plank wall is used to separate rooms. Collection locations L1 and L2 were separated by a brick wall. Collection locations 3 and 4; 5 and 6 were separated by a plank wall. The experiment was repeated every day for a week.

The degradation of signal strength for two types of walls was analyzed and shown in Fig. 3.8 and Table 3.5. The RSS values presented are the average values of all RSS samples. Different types of material have different attenuation value [98]. Fig. 3.8 obviously demonstrates that the degradation of signal strength through the brick wall (between L1 and L2) is significantly higher than the degradation strength through the plank wall (between L3 and L4; L5 and L6). Our results show that the degradation of signal strength in each pair of locations is different in every measurement. For the brick wall, the 2.4 GHz signal had a higher loss than 5 GHz signal. On the contrary, for the plank wall, 2.4 GHz signal had a smaller loss than 5 GHz signal. When transmitted through a wall, Wi-Fi signal decreases its strength, so locations between walls can be separated by different Wi-Fi fingerprints.

Table 3.4 The difference of signal strength when transmitting through a wall.

Day	L1-L2 (dB)		L3-L4 (dB)		L5-L6 (dB)	
	2.4 GHz	**5 GHz**	**2.4 GHz**	**5 GHz**	**2.4 GHz**	**5 GHz**
Day 1	15.40	10.74	-3.45	4.83	8.89	8.02
Day 2	7.16	10.17	0.88	4.14	0.15	8.12
Day 3	12.48	13.52	-1.18	4.37	4.76	8.64
Day 4	17.65	10.65	1.02	6.67	2.26	10.77
Day 5	14.27	10.54	1.40	1.73	4.31	5.93
Day 6	11.94	8.60	0.61	5.21	5.53	8.00
Day 7	18.54	14.72	4.48	8.91	0.85	9.56

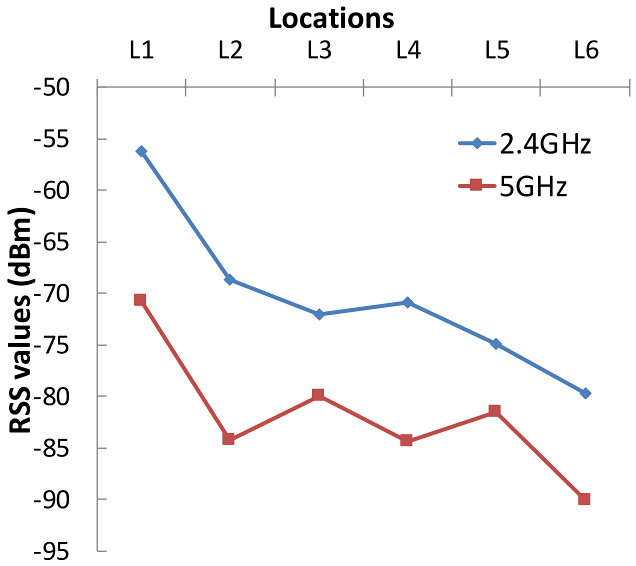


Figure 3.8 The signal strength measured between walls.

3.5 Fluctuation of Wi-Fi signal in a hall

A hall on the ground floor of an academic building is also selected as a test bed to investigate the fluctuation of the 2.4 and 5 GHz Wi-Fi signal in a hall. The dimension of the hall was approximately 20-meters long and 15-meters wide. Given APs, which work over both two spectrum bands: 2.4 GHz and 5 GHz, were placed in a line with collection locations. Smartphones were placed to measure the Wi-Fi signal at the middle of the hall with different distance to the APs, each location is one meter apart in space. The signal was measured for 30 minutes without the presence of people. Fig. 3.9 shows the layout of the experiment.

The measured data were analyzed as in corridor and office cases to calculate the average and standard deviation of a signal. Table 3.5 and Fig. 3.10 show the average value and standard deviation of the measured signal. Similar to the result measured in a corridor, at the same measurement location, the standard deviation of 2.4 GHz signal is higher than the SD of 5 GHz signal.

Table 3.5 The mean and standard deviation (SD) of signals measured in a hall without the presence of people (Unit: dBm).

Locations	**AP5_2.4GHz**		**AP5_5GHz**		**AP6_2.4GHz**		**AP6_5GHz**	
	mean	***SD***	***mean***	***SD***	***mean***	***SD***	***mean***	***SD***
M1	-69.58	1.27	-67.99	0.15	-64.96	1.02	-48.81	0.46
M2	-66.98	0.99	-71.07	0.32	-72.77	0.78	-56.11	0.31
M3	-71.40	1.22	-73.16	0.49	-70.83	0.96	-62.33	0.47
M4	-68.91	0.82	-69.02	0.17	-70.65	1.09	-62.98	0.14
M5	-65.09	0.90	-65.97	0.16	-67.77	1.47	-57.01	0.22
M6	-64.47	1.39	-66.21	0.44	-72.47	1.35	-68.13	0.48
M7	-61.42	0.93	-62.98	0.31	-79.70	1.99	-63.66	0.48
M8	-57.40	1.00	-65.00	0.46	-70.71	2.33	-65.47	0.51
M9	-49.00	0.56	-58.20	0.40	-61.04	2.70	-74.94	0.38

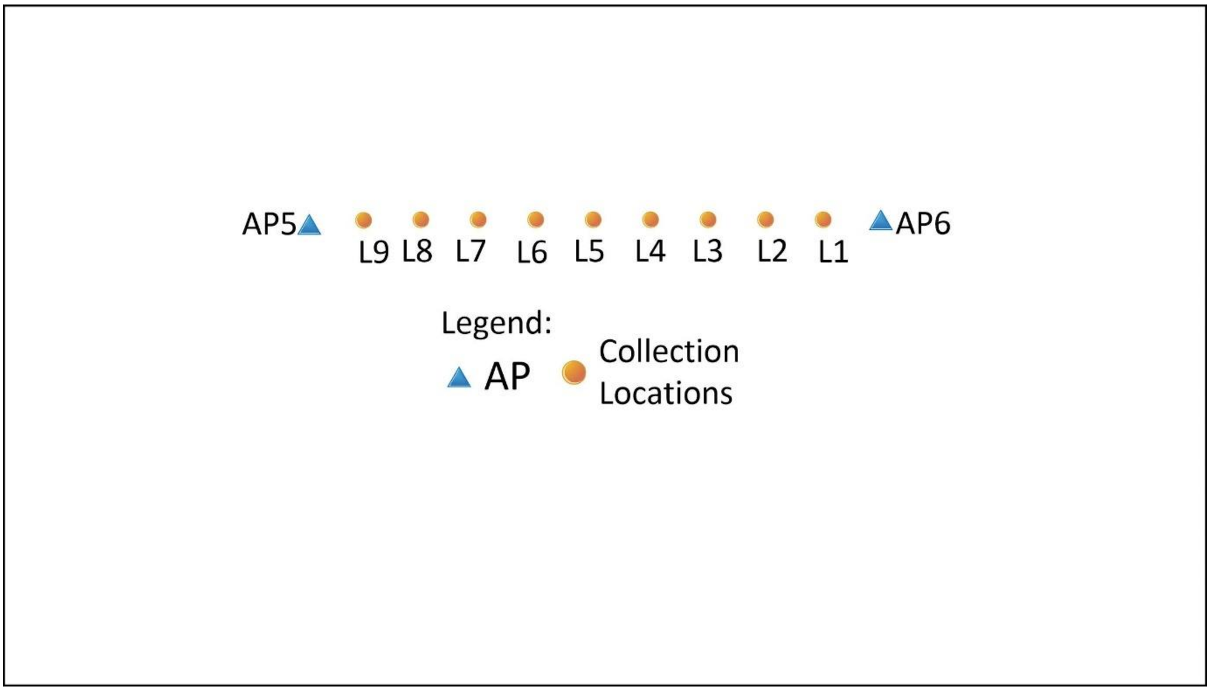


Figure 3.9 Layout of the experiment to measure Wi-Fi signal in a hall.

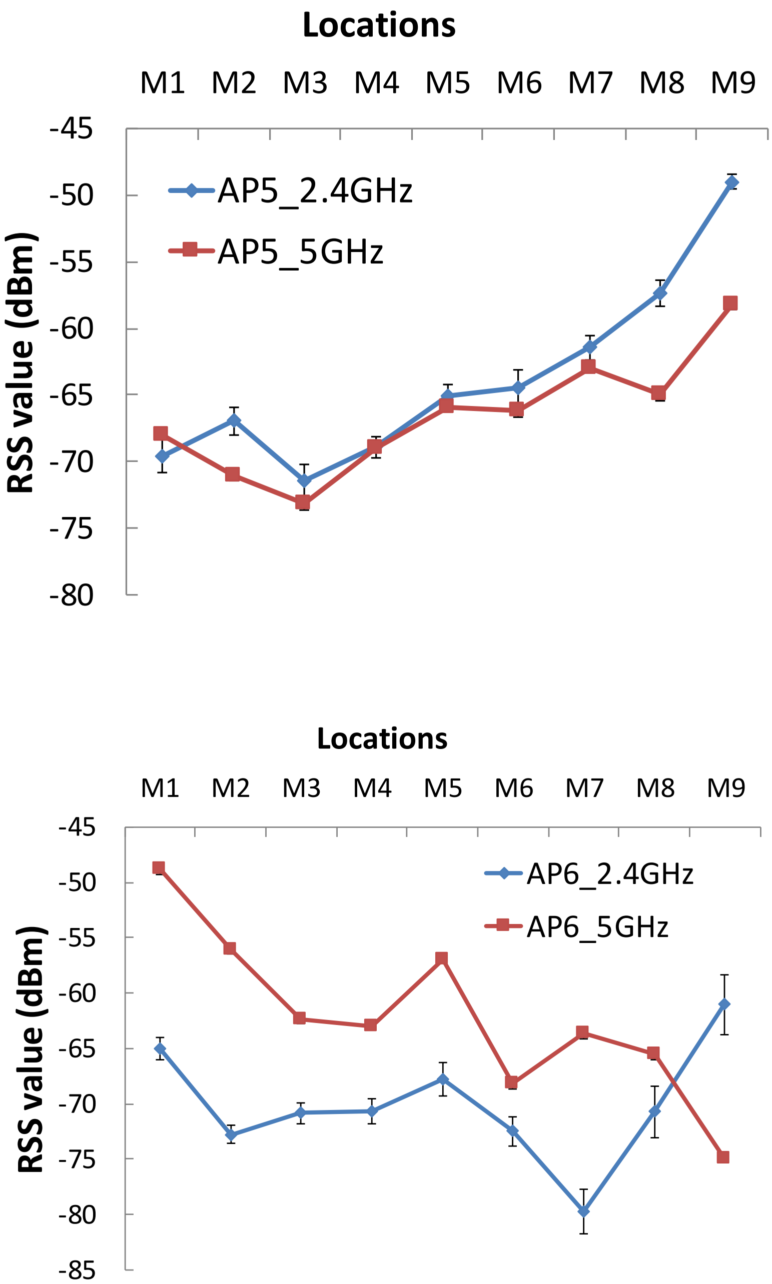


Figure 3.10 The signals measured in a hall.

3.6 Summary

In this chapter, I have investigated the fluctuation of Wi-Fi signal in a typical office environment including a corridor, hall, and rooms. The degradation of the signal through a wall was also examined. To summarize, the following findings have been presented:

- The 5 GHz Wi-Fi signal has less fluctuation than the 2.4 GHz signal in a typical office environment in all cases: rooms, corridor, and hall.
- During 30-minute measurements, the standard deviation of the 2.4 GHz signal is higher than that of the 5 GHz signal.
- The degradation of 5 GHz signal through a wall is higher than the degradation of the 2.4 GHz signal.
- At the same measurement location, the signal strength of 5 GHz signal may be stronger or weaker than that of the 2.4 GHz.
- The presence of people causes the higher fluctuation of Wi-Fi signal. The signal distributes its strength over a larger range when there are people around.
- Those observations hold true for all cases: in a corridor, a hall, and in office rooms.

4 Comparing the Performance of Wi-Fi Fingerprinting using the 2.4 GHz and 5 GHz Signals

Wi-Fi fingerprinting is a promising technique for indoor localization. The use of both 2.4 GHz and 5 GHz for Wi-Fi fingerprinting has been preliminarily studied by some research groups. However, most of the previous studies only consider the reflection and multi-paths caused by walls around but do not consider the damping and scattering through walls between them. In this study, I particularly consider the situation that the radio propagation covers several rooms divided by walls and compare the localization performance of using 2.4 GHz and 5 GHz networks. Moreover, I also compare the power consumption of scanning the surrounding wireless channels used for 2.4 GHz and 5 GHz networks. The results show that the performance of accuracy of Wi-Fi fingerprinting is similar when using 2.4 GHz and 5 GHz respectively. However, the use of 5 GHz results in higher recognition rate, while the use of 2.4 GHz consumes less power.

4.1 Introduction

Wi-Fi fingerprinting is regarded as a promising approach for indoor localization, because of its trade-off between the localization accuracy and the cost of hardware installation. The Wi-Fi networks are scanned at pre-defined locations with a mobile device and recorded as Wi-Fi fingerprints of these locations, which are stored in a fingerprint database. The phase of generating the fingerprint database is known as an off-line phase. The momentary Wi-Fi scan is compared with each of the Wi-Fi fingerprints stored in the fingerprint database, to recognize the likeliest Wi-Fi fingerprint and conclude the location accordingly.

As the 2.4 GHz band is heavily used, the less crowded 5 GHz band are emerged to avoid much of the interference at 2.4 GHz. 802.11n support the dual-band (i.e., 2.4 GHz and 5 GHz), while 802.11a and 802.11ac are operated only in the 5 GHz band [52]. The use of 5 GHz for Wi-Fi fingerprinting has been preliminarily studied by some research groups [31], [53]–[55]. They compared either the standard deviation or the statistics of received signal strength values from 2.4 GHz and 5 GHz networks, and thus, infer the potential impact on the location accuracy (i.e., the error distance) of Wi-Fi fingerprinting. In these studies, the only limited situation of radio propagation indoors is considered, such as the path loss in a hall or a large room without walls during the path.

Indeed, buildings are often organized into rooms, very often with small dimensions (e.g., an office of 10-15 m^2). The radio propagation from an access point can cover several rooms. Thus, the radio propagation in the indoor environment is complex, not only because of the reflection and multi-paths caused by the walls around, but also the damping and scattering through the walls between them. In such scenarios, it is necessary to consider the radio propagation in the area with several small-dimension rooms.

In this study, I particularly consider the situation that the radio propagation covers several rooms and compare the localization performance of using 2.4 GHz and 5 GHz networks. The rooms are divided by walls that absorb and/or reflect electromagnetic radiation and thereby contribute to making the RSS values from an AP distinguishable even in adjacent rooms [99]. In such scenarios, the location accuracy is usually defined as room-level. The room-level accuracy means the system is likely to be able to locate a mobile device in a room where the mobile device or an occupant is. I investigate the possibility of achieving room-level accuracy when using 2.4 GHz and 5 GHz networks. In addition to the analysis of RSS statistics, I also evaluate the performance of our localization system named WHERE [3], [4], which is particularly designed for room level localization (i.e., localization with room-level accuracy). The conclusion needs to be drawn from not only an analysis of the Wi-Fi signal statistics but also by considering results achieved from a localization system.

Moreover, the power consumption of scanning the surrounding wireless channels used for 2.4 GHz and 5 GHz networks is investigated. Mobile devices are used, in addition to the localization function, for many other functions, such as making a phone call, surfing the Internet, listening to music, and other entertainment. It is unpractical if the localization function consumes a large amount of power that a mobile device needs to be recharged several hours after a full charge (e.g., 4-5 hours in PlaceSense [100]). Therefore, another contribution of this study is to investigate the power consumption for localization using 2.4 GHz and 5 GHz.

4.2 Related Work

Several research groups have investigated how the underlying characteristics of radio propagation of 2.4 GHz and 5 GHz networks may affect the localization performance of Wi-Fi fingerprinting. On the one hand, the coverage of such networks is considered. The coverage distance of 5 GHz AP is smaller than the coverage distance of 2.4 GHz AP when the radio transmission powers of the two devices are equal, mainly because of the greater damping of the signals passing through walls indoors [54]. Therefore, it requires more APs of 5 GHz networks to cover the same area as for 2.4 GHz networks. On the other hand, signal stability is also investigated. For instance, Farshad et al. [31] calculate the mean and the deviation values of RSS values received from the different bands (e.g., 2.4 GHz and 5 GHz) of the same AP, and consequently present the potential impact of these bands used on Wi-Fi fingerprinting. As a result of their study, they conclude that 5 GHz networks offer relatively better location accuracy just because of its lower signal variations than 2.4 GHz. Similarly, Lui et al. [53] have investigated eleven different chipsets operating on dual bands and concluded the use of 5 GHz potentially improves the accuracy due to its higher stability than 2.4 GHz.

However, some researchers present different conclusions compared to the conclusions of the two groups above in their literature. For example, Talvitie et al. [55] have further studied the statistics of RSS values from the perspective of fingerprinting localization. They figured out that the observed RSS values of 5 GHz networks are lower than the observed RSS values of 2.4 GHz because of the larger pass loss. Suppose the high RSS values are crucial for the Wi-Fi fingerprinting, the location accuracy of the 2.4 GHz (with relatively high probability to receive high RSS values) should be better than using 5 GHz. Accordingly, their experiment results show that the localization performance with 5 GHz networks is worse than when using 2.4 GHz networks – more specifically, it results in worse accuracy and less floor wide detection probability.

The theoretical analysis and the experimental investigation of the signal coverage and stability in the literature help to understand the influence of radio frequency utilization on the performance of Wi-Fi fingerprinting. However, the results of fingerprinting localization systems utilizing 2.4 and 5 GHz are more persuasive. Besides, the use of 5 GHz networks for Wi-Fi fingerprinting does not only influence the performance of localization but may also increase the cost of resources of the mobile devices, e.g., the power consumption. One of the major obstacles to the widespread use of localization systems is the costs of the high power, compared to the limited battery life of the mobile devices [76]. PlaceSense [100] (while not focusing on battery life) has drained the battery of a smartphone within 4-5 hours.

Many research efforts have been made to reduce the power consumption. For instance, Brouwers et al. [78] have proposed an energy-efficient Wi-Fi scanning algorithm to reduce the energy consumption as only a subset of channels is scanned. Xu et al. [77] have proposed a power-save strategy for which learning and positioning occur only when the system detects the mobile devices are in a stationary state based on the built-in accelerometer sensor. Most of the approaches mentioned above are developed in the scenario of 2.4 GHz networks. As one of the performance measures for the system reliability in practice, the power consumption demanded when running the fingerprinting localization systems in the scenarios of 2.4 GHz and/or 5 GHz networks needs to be well-considered.

In previous works, Kaemarungsi et al. [23] have listed the underlying factors that may affect the localization performance of Wi-Fi fingerprinting approaches. In this chapter, I compare the performance of Wi-Fi fingerprinting when using 2.4 GHz and 5 GHz networks, regarding the distribution, mean, and standard deviation of RSS values. In addition to the analysis of the signals, I also compare the Wi-Fi fingerprints learned with a Wi-Fi fingerprint system WHERE [3], [4] in the real-world scenarios. Furthermore, the power consumption of running this fingerprinting system is also presented.

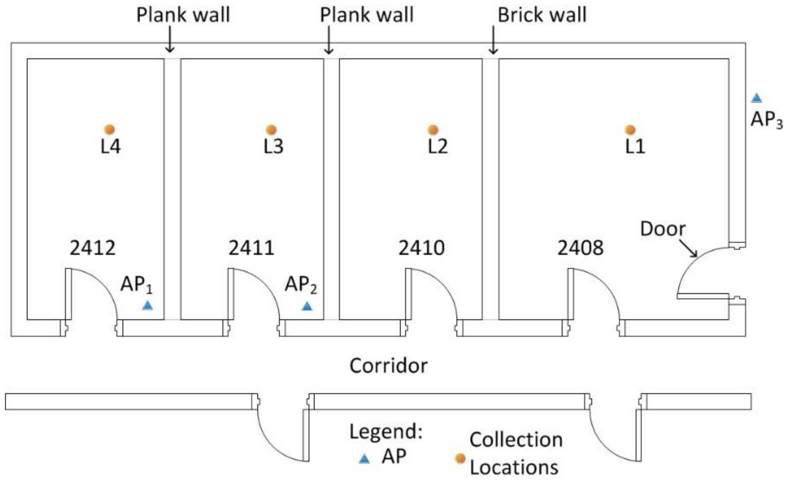


Figure 4.1 The location of the given APs and the data collection locations.

4.3 Experimental Investigation

The research was conducted in an office area with four office rooms next to each other. The layout of these rooms is shown in Fig. 4.1. In this study, all APs were set up to broadcast both 2.4 GHz and 5 GHz frequency bands simultaneously. Four mobile phones (Nexus 5) were used as measurement devices with the Wi-Fi sampling interval in five seconds and were placed in a fix location at the height of a table (about 70cm high) to measure the Wi-Fi RSS from the surrounding APs. The signals were measured for 30 minutes without the presence of people. The experiment was repeated every day for one week.

The fluctuation of the signals received from 2.4 GHz and 5 GHz networks was investigated by generating the signal histogram and calculating the standard deviation of the signals. Our fingerprinting system WHERE [3], [4] is used to test the performance of a Wi-Fi fingerprinting system using 2.4 GHz and 5 GHz signals. In the learning phase, the system WHERE applies the density-based clustering algorithm [24] to construct a fingerprint database of referent points, each fingerprint consisting of a set of RSS-range. In the positioning phase, the system compares the momentary Wi-Fi scans with the fingerprints in the fingerprint database and returns the fingerprint associated with the current location. In this study, the Wi-Fi data collected on the first day are used as training data to construct a fingerprinting database. Then, the Wi-Fi data measured in the other days are used as test data to evaluate the performance (e.g., accuracy, recognition rate) of the system. For the evaluation, the output of the system can be *true* or *false*.

- True: the mobile device is in a location / room, and the system returns a fingerprint associated with the location / room.

- False: the mobile device is in a location / room, but the system returns a fingerprint associated with a different location / room.

- Non_Response: the mobile device is in a location / room, but the system does not return a positioning response.

The performance of the system is evaluated using the metric of *accuracy*, which is defined as:

- $\text{Accuracy} = \frac{\text{True}}{\text{True+False}}$

The metric of *recognition rate* indicates how often the system answers the location requests. A system which has a higher recognition rate can reduce the time needed for positioning, as well as reducing the power consumption. The metric of recognition rate is defined as:

- $\text{Recognition rate} = 100\% - \frac{\text{non\_response}}{\text{total testing}}$

The power consumption of performing the Wi-Fi scanning activity is also investigated. I use two multimeter devices, Peaktech 3415 USB digital, to measure the voltage and current of a smartphone while scanning 2.4 GHz and 5 GHz bands. Those two multimeters are also connected to a computer to record the measurement values with a sampling rate of 2 Hz as described in [101]. Then the average power consumption was calculated. All other sensors of mobile phones are turned off.

4.4 Results

4.4.1 The fluctuation of 2.4 GHz and 5 GHz Wi-Fi signals

The Wi-Fi signal varies over time [63]. This fluctuation causes the difference of the RSS in different measurements and may influence the performance of a Wi-Fi fingerprinting system. In this section, I compare the fluctuation of Wi-Fi signals from different Wi-Fi frequency bands (i.e., 2.4 GHz and 5 GHz) of an AP. Fig. 4.2 shows that the histogram distribution of samples from 5 GHz tends to concentrate in a smaller range, whereas samples from 2.4 GHz distribute its values over a larger range. In addition, the standard deviation of 5 GHz Wi-Fi signals is smaller than the SD of 2.4 GHz signals (Table 4.1) in most cases. The result here is similar as results shown in some other literature [31], [54]. However, when the distance of the transmitted signals increases, the Wi-Fi signals from 5 GHz becomes weak and unstable. For instance, the SD of the AP3_5GHz is remarkably higher than the SD of 2.4 GHz signals (5.65 dBm versus 0.55 dBm for location 2, and 8.36 dBm versus 1.62 dBm for location 3).

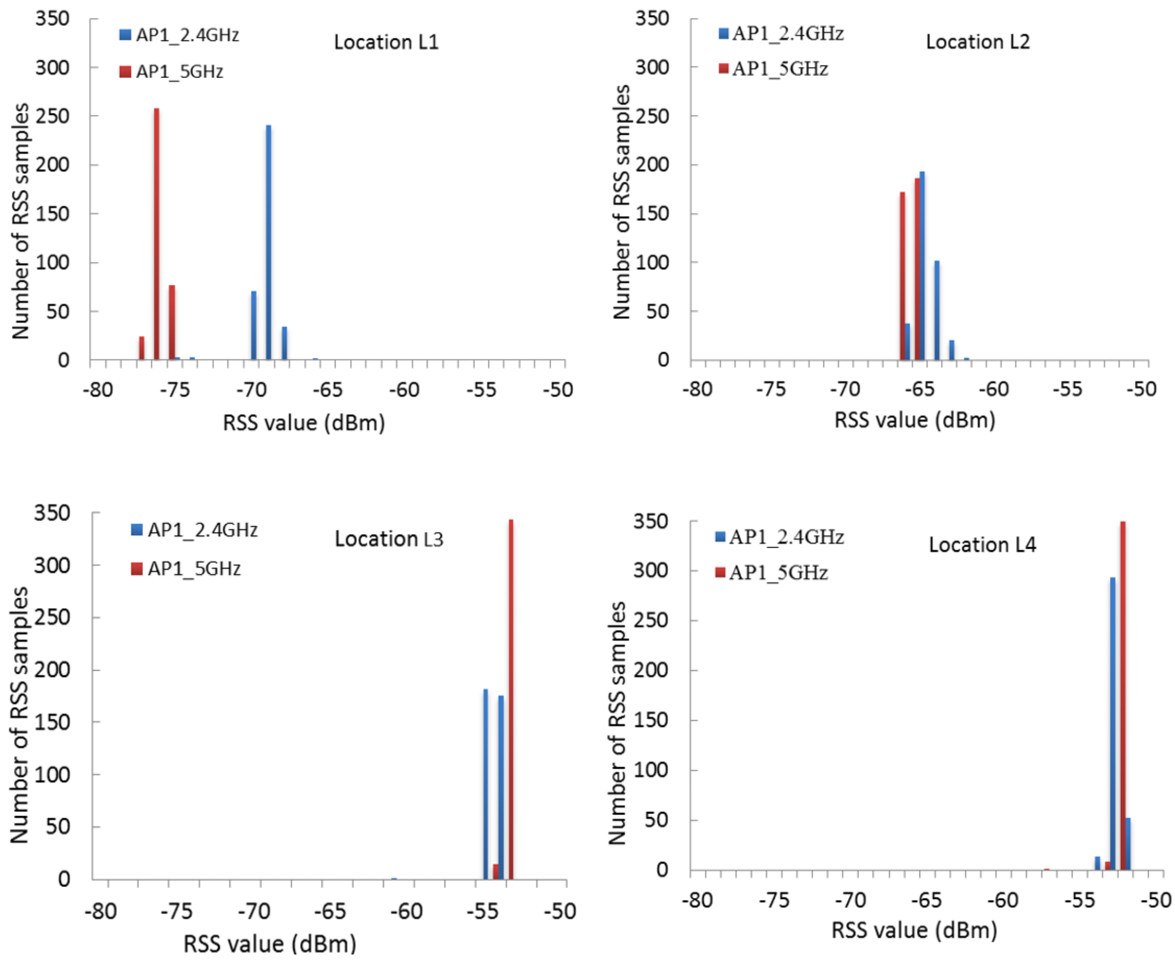


Figure 4.2 The histogram of 2.4 and 5 GHz signals.

4.4.2 The fingerprint range of 2.4 and 5 GHz Wi-Fi signals

The fingerprint of a location in the system WHERE consists of a set of high-density RSS-ranges [102]. On the one hand, if the RSS-range of a location is small, the accuracy is usually high but the recognition rate is usually low. On the other hand, if the RSS-ranges of two location nearby are overlapping, it is impossible to distinguish these two locations, and thus, the accuracy is low but the recognition rate is usually high. Hence, I also investigate the RSS-ranges generated from the Wi-Fi data using 2.4 GHz and 5 GHz. In Fig. 4.3, the dashed blue line represents the RSS-ranges of 2.4 GHz signals, while the unbroken red line is the RSS-ranges of 5 GHz signals. Similar to the standard deviation, the fingerprint of 5 GHz signals also has a smaller RSS-range than that of 2.4 GHz signals. The larger fingerprint range of 2.4 GHz signals seems to help the test samples easier to be matched with the fingerprint of a location. However, the larger fingerprint range may also increase the probability of mismatching the test samples to the wrong locations. On the contrary, the smaller fingerprint range of 5 GHz signals may help to reduce the mismatch, increase the accuracy, and help to distinguish locations in a smaller distance.

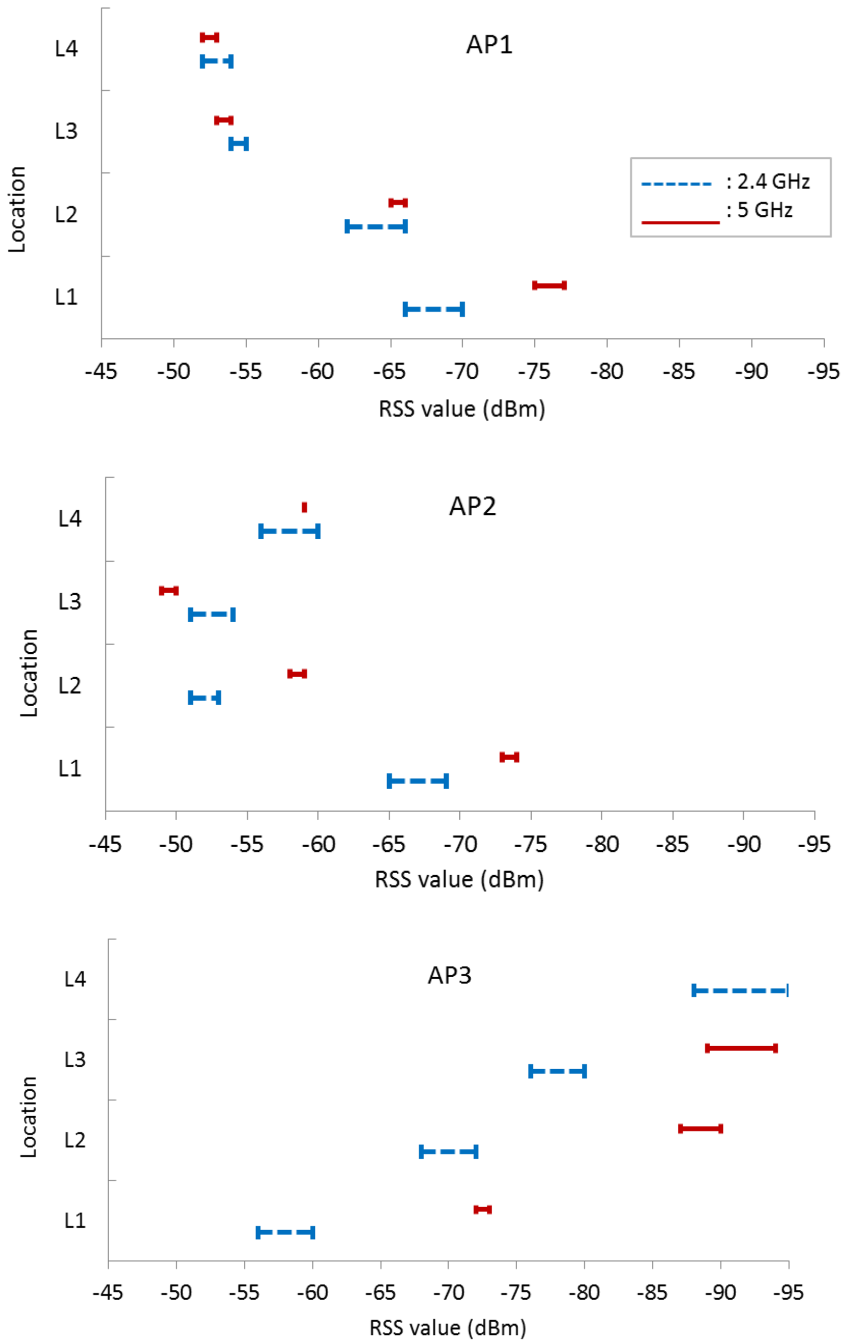


Figure 4.3 The RSS fingerprint range of the 2.4 and 5 GHz bands.

4.4.3 The performance of WHERE using 2.4 and 5 GHz signals

The above analyses show that the fluctuation and the RSS-ranges of 2.4 and 5 GHz signals are different. Therefore, the performance of a fingerprinting system using Wi-Fi data from those different frequency bands may also be different. This analysis investigates the performance of our Wi-Fi fingerprinting system WHERE under three cases. The system consecutively generated the fingerprint database using data measured from only 2.4 GHz band, only 5 GHz band, and both bands. Unlike the assumption that the performance of Wi-Fi fingerprinting is influenced by the signal frequency, the result in Table 4.2 shows that we can achieve similar accuracy result in using the 2.4 GHz, 5 GHz, and both bands. The overall correct classifying result of using 2.4 GHz only, 5 GHz only, and both bands is 99.95%, 99.98%, and 100%, respectively. Although the accuracy results are similar, the recognition rate using the 5 GHz signals is considerably higher than the recognition rate of using 2.4 GHz signals (Table 4.3). The explanation for the higher recognition rate of the 5 GHz signals may be because of the smaller fluctuation of the RSS at 5 GHz. Thus, a localization system using 5 GHz signals achieves faster response than using 2.4 GHz signals. The recognition rate of using both band signals is not much higher than that of using 5 GHz signals.

Table 4.1 The mean and standard deviation (SD) of signals in different locations (Unit: dBm).

Access Point	**L1**		**L2**		**L3**		**L4**	
	mean	***SD***	***mean***	***SD***	***mean***	***SD***	***mean***	***SD***
AP1_2.4GHz	-69.15	1.11	-64.70	0.89	-54.53	0.61	-52.89	0.41
AP1_5GHz	-75.85	0.51	-65.48	0.50	-53.04	0.20	-52.04	0.30
AP2_2.4GHz	-67.82	1.52	-51.77	0.64	-52.97	2.04	-58.53	0.60
AP2_5GHz	-73.02	0.14	-58.23	0.42	-49.95	0.33	-59.00	0.00
AP3_2.4GHz	-59.00	0.76	-71.48	0.55	-79.26	1.62	-91.84	1.08
AP3_5GHz	-72.04	0.41	-87.35	5.65	-89.24	8.36		

Table 4.2 The accuracy (%) of using different frequency bands and APs.

	L1	**L2**	**L3**	**L4**	**overall**
3APs _2.4GHz	100	100	100	99.42	99.95
3APs _5GHz	100	100	100	99.92	99.98
Both bands	100	100	100	100	100

Table 4.3 The recognition rate (%) of using different frequencies and APs.

	L1	**L2**	**L3**	**L4**	**overall**
3APs _2.4GHz	18.55	19.40	40.42	6.96	21.59
3APs _5GHz	71.31	63.04	26.48	53.06	52.11
Both bands	63.05	66.02	54.32	38.55	54.97

4.4.4 The power consumption of scanning 2.4 GHz and 5 GHz signals

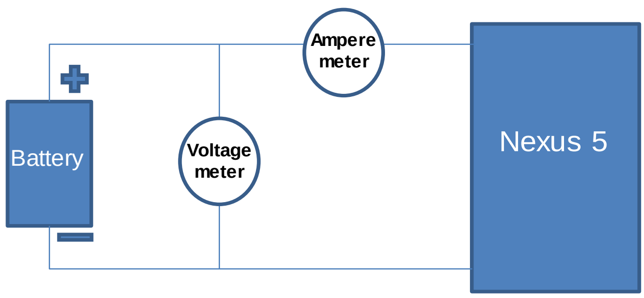


Figure 4.4 Connecting Voltmeter and Ammeter to measure the power consumption.

This analysis investigates the power consumption to perform a Wi-Fi scanning in three cases: scanning only the 2.4 GHz band, only the 5 GHz band, and both bands. Every test was measured for 15 minutes. Fig. 4.4 shows how the Voltmeter and Ammeter are connected to the mobile phone to measure the power consumption.

- First, the power consumption of a mobile phone in the idle state (did not run Wi-Fi scanning) was measured.
- Second, the power consumption of a mobile phone when performing Wi-Fi scanning task in each frequency band with different scanning intervals in 5, 10, and 20 seconds was measured.
- Then, the difference between the first and the second measurement was calculated to find out the power consumption to perform the scanning.

The results in Fig. 4.5 indicate that scanning more Wi-Fi channels consumes more energy. With the same scanning interval, the energy consumption of scanning both bands is higher than the energy consumption of scanning only 5 GHz band or only 2.4 GHz band. Scanning 5 GHz band consumes more power than scanning 2.4 GHz band due to the higher number of channel of 5 GHz band. By increasing the scanning interval from 5 seconds to 10 or 20 seconds, the power consumption reduces in all three cases. For examples, for the case of scanning both bands, the power consumption decreases from 210 mW to 173 mW and 131 mW with the scanning intervals being 5 seconds, 10 seconds, and 20 seconds, respectively. That information provided valuable insights into the power consumption of scanning Wi-Fi signal. In section 4.4.3, I show that the recognition rate of using 5 GHz band in our system is higher than the recognition rate of using 2.4 GHz band while the accuracy of the system using 2.4 GHz and using 5 GHz is similar. These results suggest that a Wi-Fi fingerprinting system may use

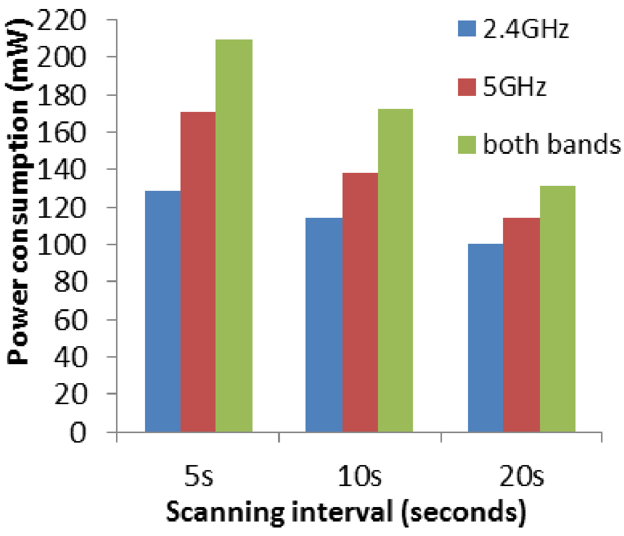


Figure 4.5 The power consumption of performing Wi-Fi scanning.

the 5GHz signals when requesting higher recognition rate, whereas using the 2.4 GHz signals when requesting low power consumption.

4.5 Summary

This chapter has particularly investigated the performance of Wi-Fi fingerprinting in respect to the signal fluctuation, the fingerprint range, the recognition rate, the power consumption, and the accuracy in the scenario of indoor space with several small rooms divided by walls, which is a quite typical office area inside buildings. The use of Wi-Fi signals received from 2.4 GHz networks only, from 5 GHz only, and from both bands are compared. The results show that the performance of accuracy of Wi-Fi fingerprinting is similar when using 2.4 GHz and 5 GHz bands. The performance of the recognition rate of a system using signals of 5 GHz was found to be better than that using 2.4 GHz signals because both the fluctuation and the fingerprint ranges generated from 5 GHz signals are smaller than that of generating from the 2.4 GHz signals. Besides, scanning the 2.4 GHz networks consumes less power than scanning the 5 GHz networks. The results indicate that the selection of scanning the frequency bands in Wi-Fi fingerprinting approaches may not influence the results of accuracy but influence the recognition rate and the power consumption. The trade-off of the performance needs to be carefully considered when designing an indoor localization system when using Wi-Fi fingerprinting.

5 The Influence of Small-scale Fading and Device Heterogeneity on Wi-Fi Fingerprinting

As Wi-Fi fingerprinting is a promising approach for indoor localization, one of the major challenges is understanding how small-scale fading and device heterogeneity problems influence the performance of the Wi-Fi fingerprinting systems. In this chapter, I experimentally investigate the influence of small-scale fading and the device heterogeneity problems on the Wi-Fi RSS values, as well as on the performance (i.e., the recognition rate) of a Wi-Fi fingerprinting system WHERE. Besides, I also compare the performance of Wi-Fi fingerprinting in the experimental scenarios and in the real scenarios with the consideration of the small-scale fading problem. It concludes that both small-scale fading and device heterogeneity heavily influences the RSS values. However, the influence of small-scale fading on the performance of the Wi-Fi fingerprinting system can be mitigated when the Wi-Fi data is collected in the real scenarios.

5.1 Introduction

Wi-Fi fingerprinting has become a promising approach to locate the user/mobile device indoors, because Wi-Fi networks are already installed in many buildings. It usually works in two phases: the training phase and the positioning phase. In the training phase, the Wi-Fi networks are scanned to generate the fingerprint of reference places. These fingerprints are stored in the fingerprinting database. In the positioning phase, the momentary scan of Wi-Fi signals is compared with the fingerprint stored in the fingerprinting database to figure out the user's current position. The critical assumption made in the use of Wi-Fi fingerprinting is that the fingerprint are unique for each location, and they do not vary over time. However, the received signal strength values measured with mobile devices (e.g., smartphones) and the performance (e.g., accuracy, recognition rate) of Wi-Fi fingerprinting are influenced by the surrounding Wi-Fi environment and the devices used for localization.

In particular, the Wi-Fi network protocols and chips have not been designed for localization. Therefore, a major issue is the device heterogeneity (or known as device diversity) problem. The training devices used to generate the fingerprints during the training phase may differ from the positioning devices used during the positioning phase. Different devices with different Wi-Fi chipsets and antennas measure varying RSS values, even though the devices are placed at the same location [53], [63]. Some publications report that the difference among different devices may be up to 30 dB [53]. Varying RSS values from different devices can influence the performance of Wi-Fi fingerprinting. Besides, another major issue for Wi-Fi fingerprinting is small-scale fading which results in the severe fluctuation of the RSS values when the devices move over a very short distance of half a wavelength of the Wi-Fi signals. As a consequence, the fluctuation of RSS values may also influence the performance of Wi-Fi fingerprinting. The influence of small-scale fading on the RSS values has been investigated in the literature [56], [57]. However, how small-scale fading influences the performance of a Wi-Fi fingerprinting system has not been well-investigated yet. For instances, the Wi-Fi fingerprints of a Wi-Fi fingerprinting localization system WHERE is a set of high-density RSS-ranges instead of a set of RSS values. The influence on the RSS-ranges, as well as on the performance of such a system is not addressed.

This chapter makes the following contributions:

- Using a statistical test to validate the influence of small-scale fading and the device heterogeneity on the Wi-Fi RSS values.
- Investigating the performance of a Wi-Fi fingerprinting system under the influence of small-scale fading and the device heterogeneity in the experimental scenarios.
- Comparing the performance of the Wi-Fi fingerprinting system in the experimental scenarios and in the real scenarios, and thus, proposing a method to mitigate the influence of small-scale fading with the assistance of the imbedded accelerometer sensors.

5.2 Related Works

The influence of device heterogeneity on Wi-Fi signals has been studied in some literature. The authors in [53], [63] report that different Wi-Fi chipsets from different manufacturers perform differently, and therefore give different RSS values even though the devices are used to measure the signals at the same location. This may influence the performance of a Wi-Fi fingerprinting system if various devices are used in the training phase and in the positioning phase. Several methods have been introduced to address the problem of using heterogeneous devices in a Wi-Fi fingerprinting localization system. One popular method is to manually calibrate the fingerprints based on the linear transformation [70]. Other approaches [42]–[44], [73] extract features (e.g., differences of RSS values or the ration of RSS values between pairs of APs), instead of using RSS values, in order to eliminate the influence of device heterogeneity. Wang et al. [45] proposed a spatial mean normalization (SMN) method to address the variation in heterogeneous hardware. The SMN method mitigates the difference of the antenna gain among heterogeneous devices by calculating the difference between the absolute RSS and the spatial mean RSS values of the observed APs. Another approach [46] called rank-based fingerprinting uses the rank of the APs as fingerprints.

The influence of small-scale fading on Wi-Fi signals has also been studied. In [57], the authors investigate the susceptibility to the fading effect of the Bluetooth signal. They reported that there is a deep fading in the Bluetooth signal strength when the measurement position is moved even just 10 cm. A method to mitigate the fading is proposed using the window method with max, mean, and median value. V. Moghtadaiee and A. G. Dempster [56] investigate the relationship between the geometric distance and the vector distance by analyzing data measured at a very short distance at [0, 1, 5, 10, 20, 50, 100, 200] cm. The authors concluded that the RSS variation due to the small-scale fading effect is significant even when the geometric distance of the measurement positions is very small. These papers conclude the performance (i.e., accuracy) of Wi-Fi fingerprinting is influenced with small-scale fading because of the variation of the RSS values. However, when the Wi-Fi fingerprints are more than a set of RSS values, the influence of small-scale fading and device heterogeneity on the performance of the Wi-Fi fingerprint systems are not well-investigated yet.

5.3 Experimental Methods

In this study here, all APs are two frequency band devices and were set up to broadcast both 2.4 GHz and 5 GHz frequency bands simultaneously. Five mobile phones (Nexus 5) were used as measurement devices and were placed on a table to measure the Wi-Fi RSS from surrounding APs. The measurement devices were placed in the same direction in all measurements. Every measurement was performed in 5 minutes without the presence of people in order to provide a stable radio environment.

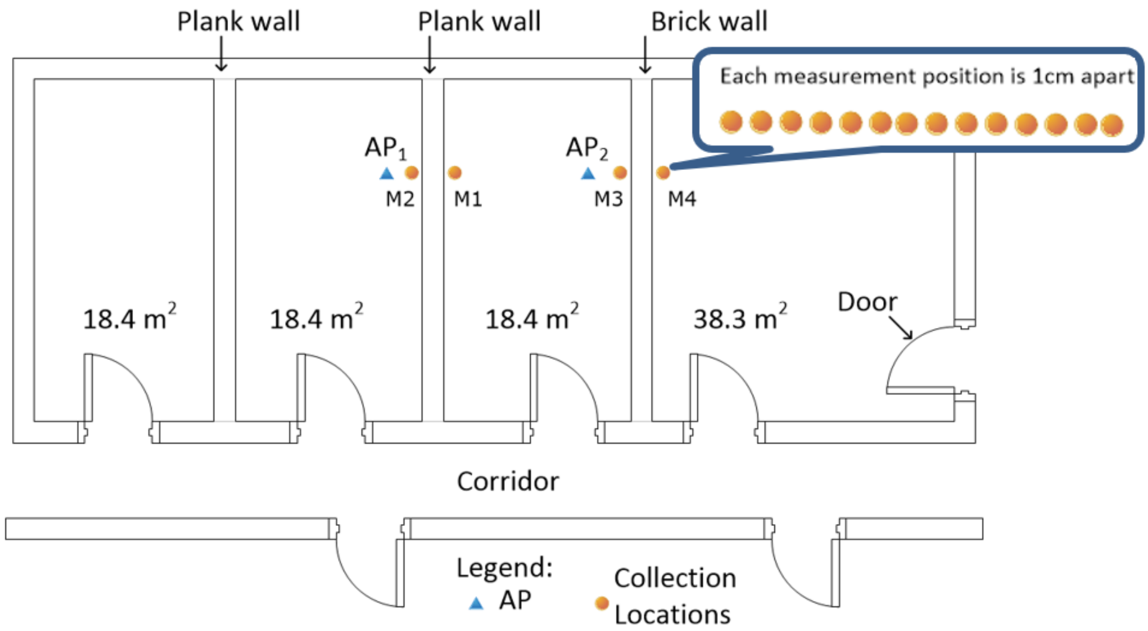


Figure 5.1 The positions of APs and measurement devices for the small-scale fading experiment.

To investigate the difference of Wi-Fi RSS measured in different positions and by different measurement devices, the data analysis tool of MATLAB is used to generate the box-and-whisker plot, to perform the Analysis of Variance test and the t-test [91].

The box-and-whisker plot is used to describe the median, min, max value, the lower quartile, and the upper quartile of a set of Wi-Fi RSS data. Quartiles divide the data set into four equal parts; each part shows 25% of the data. The central box covers the first to the third quartile to show the middle 50% of data points. A line inside the box indicates the sample median. If the box appears as a single line, it indicates most of the data contains the same single value for that group. The whiskers extend from the quartiles to the smallest and largest data values in each sample. Unusual points far from the box are indicated by a plus sign.

ANOVA is a popular statistical test used to compare the difference of a group mean. In the ANOVA test, the null hypothesis of equal population means of all samples is tested to see whether it is true or not. The t-test is used to test the null hypothesis that two sets of data have equal mean. The RSS values measured from a specific AP follow a Gaussian distribution [71], so the t-test is a good test to compare two sets of RSS values. The result of the ANOVA test and the t-test is considered statistical significant if the p-value is less than or equal to the significant level $\alpha = 0.05$ which indicates a 5% risk to conclude that there is a difference among group means [87]–[89], [91].

To investigate the influence of the fading effect and the heterogeneous device problem to the performance of a Wi-Fi fingerprinting system, our fingerprinting system WHERE [3], [4] is used. WHERE applies the density-based clustering algorithm [24] to construct a fingerprinting database from the collected Wi-Fi data. Then, in the testing phase, the system compares the Wi-Fi testing samples with the fingerprint-range to see whether the samples can be recognized or not. If the Wi-Fi RSS of the test samples falls into the range of the fingerprint, those test samples are recognized. Otherwise, they are considered as non-recognized samples. In this study, the recognition rate was used as a metric to investigate the performance of our system under the influence of the fading effect and the heterogeneous devices. The analysis tool WEKA [103] is used to generate data to perform five-fold cross-validation.

5.4 Results

5.4.1 Variation of Wi-Fi RSS values due to small-scale fading

This experiment investigates the influence of small-scale fading on the Wi-Fi RSS values in an indoor environment. The measurement is done in an office area as shown in Fig. 5.1. Each of the measurement devices (M1, M2, M3, and M4) measures the RSS values in 14 measurement positions, respectively. One position is one centimeter apart from another in space. In other words, the distance between the AP and the mobile is increased 1 cm in the consecutive measurements. The distance between the first and the last position is 13 cm which is longer than the wavelength of the 2.4 and 5 GHz signals.

The measurement devices M1, M4, AP2_2.4GHz and AP2_5GHz were selected as examples to show the box-and-whisker plot of Wi-Fi RSS. Fig. 5.2 shows the RSS values that are measured by the same device at the 14 measurement positions. The maximum difference of RSS values measured with M1 to AP2_2.4GHz is 8 dB, while the maximum difference to AP2_5GHz is 7 dB. The maximum difference RSS measured with M4 to AP2_2.4GHz is 9 dB, while the maximum difference to AP2_5GHz is 10 dB. It can be seen that the RSS values severely fluctuate, although the distance between the measurement positions is just centimeters.

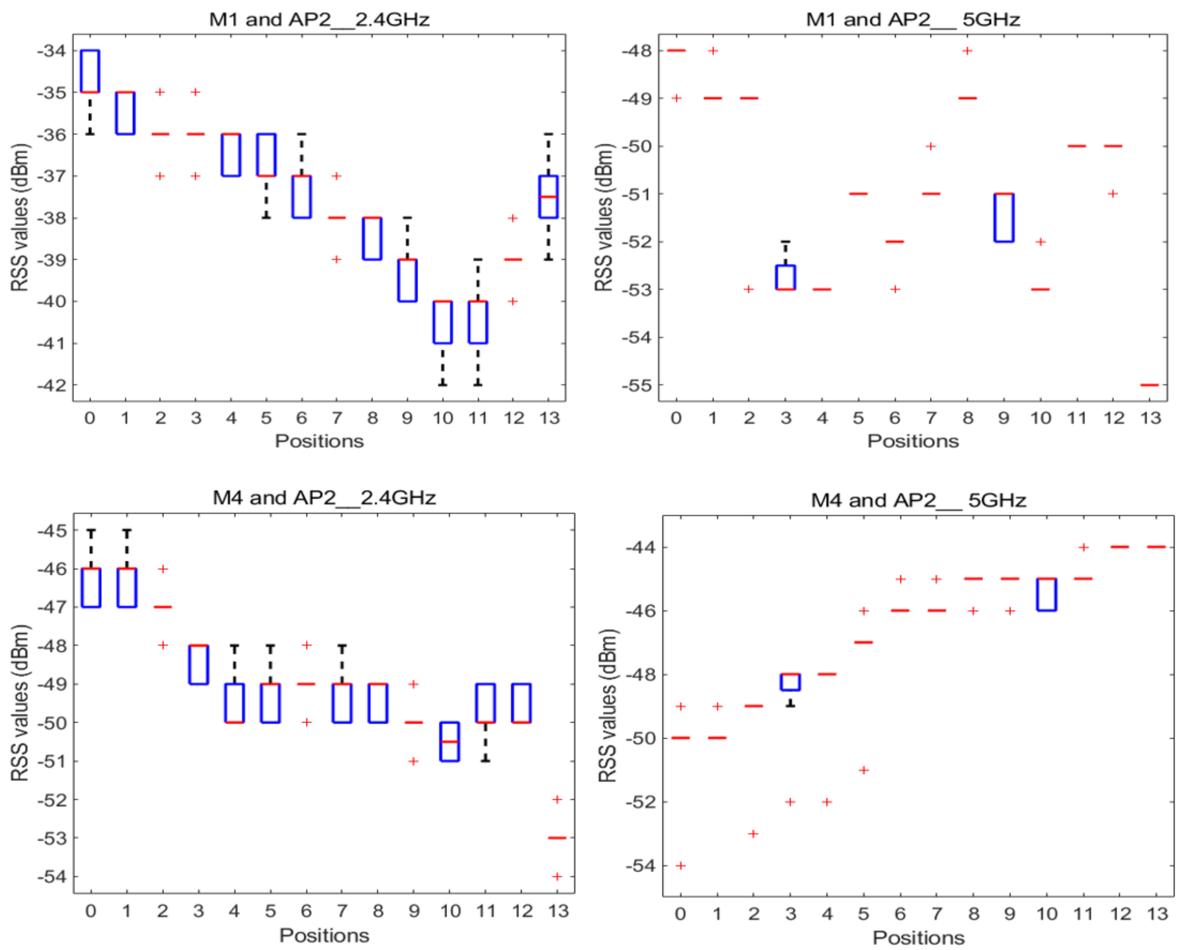


Figure 5.2 The Wi-Fi signal variation over short distant positions, each is one cm further away than the previous position.

To verify the observation, the ANOVA test was used to evaluate the hypothesis that the mean RSS of 14 measurement positions is equal. The statistical result of ANOVA test with the p-value < 0.05 indicates that the means of 14 groups are not all equal with the significant level of 05%. To have a more specific result of which pair of data is similar or different, the t-test of all pairs of data was performed. From 14 measurement positions of each device, 91 pairs of any two positions were generated to perform the t-test to see whether that pair is different or not. The t-test results show that many pairs have different mean RSS (Table 5.1).

Table 5.1 Number of pairs of data show different mean due to small-scale fading (p-value < 0.05).

APs	**Measurement devices**			
	M1	**M2**	**M3**	**M4**
AP1 _2.4GHz	88	72	90	89
AP1 _5GHz	85	87	84	79
AP2 _2.4GHz	89	88	90	84
AP2 _5GHz	86	86	78	80

5.4.2 Variation of Wi-Fi signals due to device heterogeneity

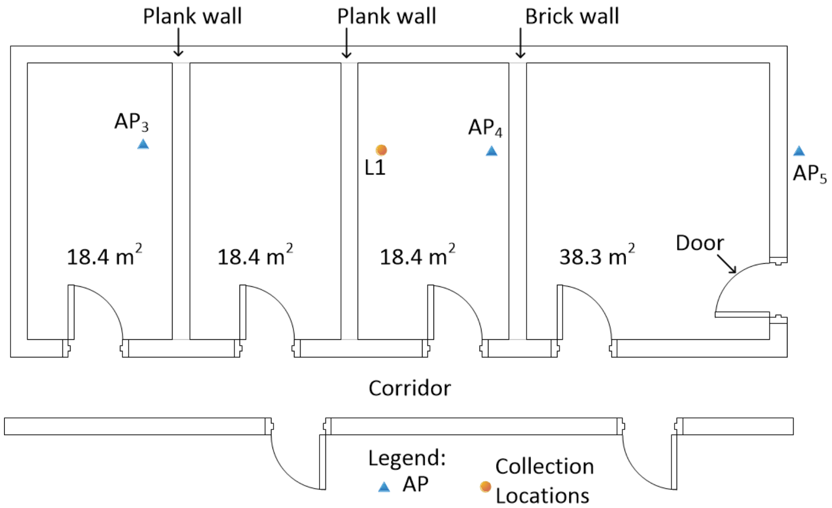


Figure 5.3 The positions of the given APs and measurement devices for the device heterogeneity experiment.

This experiment aims to compare the Wi-Fi signal measured by different devices at the same location. In this experiment, five measurement devices were placed in location L1, one after another, to measure the Wi-Fi signal (Fig. 5.3). The location to place the measurement device is marked carefully to make sure all measurement devices will be placed at the exact area. Fig. 5.4 demonstrates the Wi-Fi signal measured by five different devices at the same location L1. It can be observed that the RSS values are not similar even though they are measured at the same location. To examine the similarity of the RSS measured by different devices, I also run the statistical ANOVA test to test the hypothesis that the mean RSS of those device data sets is equal. The result of the ANOVA test proves that there is a statistically significant difference between the mean of the data sets. The t-test was then used to test the mean value of the data set measured by two different mobile devices. For five measurement devices in location L1, ten pairs of mobile devices are established to perform the t-test. Consequently, the t-test result shows that many pairs have different mean. Table 5.2 shows the number of pairs of measurement devices has significantly different mean according to each location and AP. This experiment result proves the heterogeneous device problem when using different measurement devices to measure Wi-Fi signal.

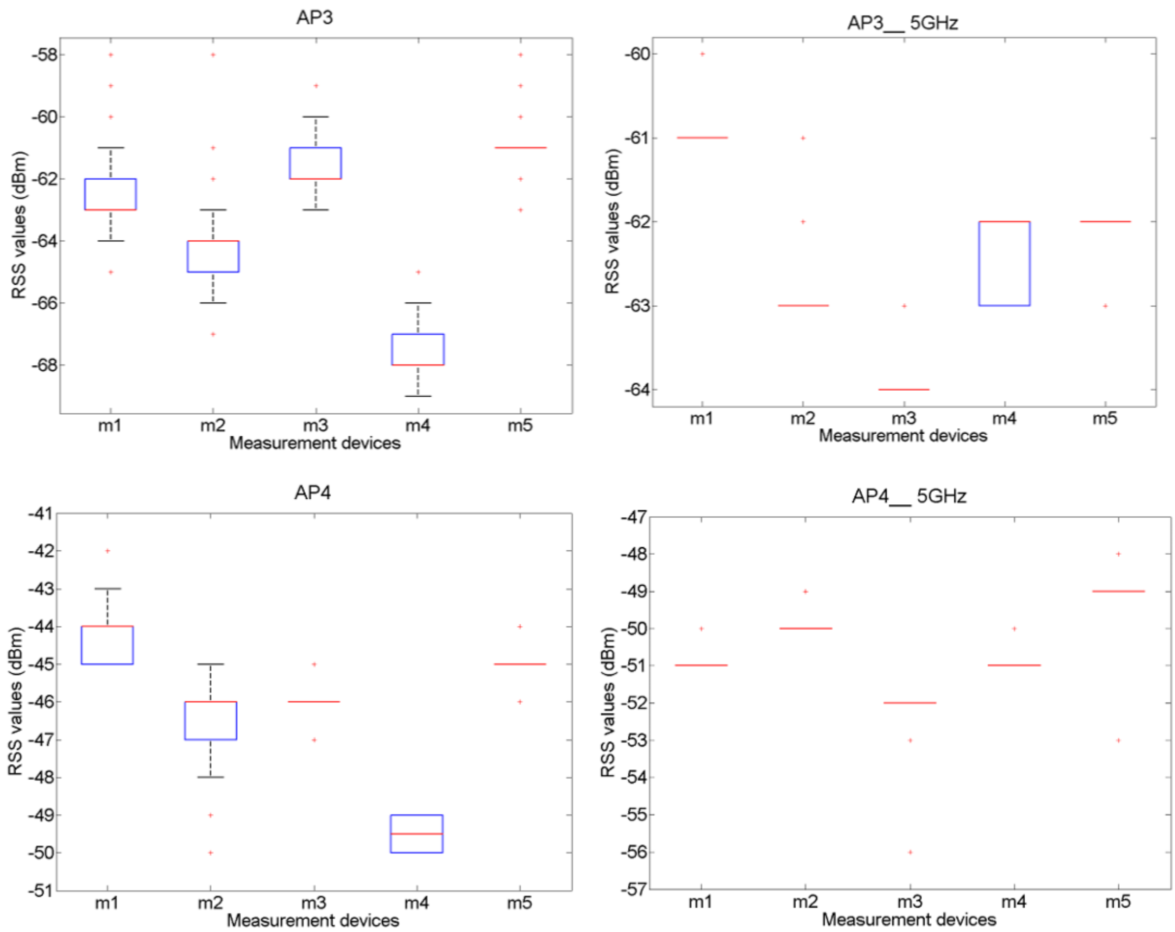


Figure 5.4 The Wi-Fi signal measured by different devices at the same location.

Table 5.2 Number of pairs of data show different mean due to device heterogeneity (p-value < 0.05).

APs	Location 1
AP3 _2.4GHz	10
AP3 _5GHz	10
AP4 _2.4GHz	9
AP4 _5GHz	9
AP5 _2.4GHz	8
AP5 _5GHz	9

5.4.3 The performance of the Wi-Fi fingerprinting system in the experimental scenarios with the consideration of the influence of small-scale fading and device heterogeneity

The recognition rate shows how often the location can be recognized when the momentary scans are compared with the fingerprint database in the positioning phase. The dataset described in section 5.4.2 are used to investigate how small-scale fading and device heterogeneity influence the recognition rate. The Wi-Fi fingerprinting localization system WHERE is used for the investigation. The datasets measured by each of the five devices are used to generate fingerprints of each device at location L1.

- First of all, I compared the recognition rate when the same measurement devices are used in both the training and the testing phase. I divided five-minute data into three-minute and two-minute data for training and testing, respectively. By using the same measurement devices at the same location, I suppose that the results are not influenced by small-scale fading and device heterogeneity. The recognition rate shows that 100% of test data can be recognized (Table 5.3).
- Secondly, I use different devices as the training device and the positioning device. This analysis reveals the influence of the heterogeneous devices on the performance of the fingerprinting system. The recognition rate shows that 47.76% of test data can be recognized.
- Thirdly, I investigate the influence of small-scale fading on the recognition rate when the measurement positions (L1, L2, L3, L4, and L5) are centimeters apart. Dataset collected

in L1 are used as training data to generate the fingerprint database, and the datasets collected in L2, L3, L4, and L5 are used as test datasets. The recognition rate shows that the locations nearby cannot be recognized due to small-scale fading.

Table 5.3 The recognition rate under the influence of the fading effect and the device heterogeneity.

Training phase versus Testing phase	**Total test samples**	**Recognition rate (%)**
use same device, at same location	246	100
use different devices, at same location	2460	47.76
use same device, at different locations	465	0

5.4.4 The performance of the Wi-Fi fingerprinting system in the real scenarios with the consideration of small-scale fading

The analysis results in section 5.4.3 have shown the severe influence of small-scale fading on the Wi-Fi RSS values, even when the measurement positions are only centimeters apart. The recognition rate of a Wi-Fi fingerprinting system is poor due to the fading effect. In fact, the mobile devices carried by users are nearly impossible to be fixed exactly at a measurement position in the real scenarios, since people always move around or shake the body a bit as time goes. Therefore, the devices carried by users in the real scenarios can collect a set of data when the measurement positions change over a small distance. Here, the collection of Wi-Fi data measured in L1, L2, L3, L4, and L5 was considered as the training and test dataset in the real scenarios and generate the fingerprints. The system WHERE generates the fingerprint using the training dataset and calculates the recognition rate using the test dataset. In such scenarios, the recognition rate, in this case, is 92.13%, as shown in Table 5.4. It can be seen that the use of Wi-Fi data collected in the real scenarios can mitigate the influence of small-scale fading on the performance of the recognition rate, since the training data is a collection of Wi-Fi data collected over a short distance in space.

Table 5.4 Solving the influence of the small-scale fading.

Solution	**Total test samples**	**Recognition rate (%)**
Combine data measured at different positions	2073	92.13

Based on these results, I propose to use the accelerometer sensor as an "assistant" to improve the reliability of the fingerprint database. The accelerometer sensor can be used to detect the motion state of the measurement device. On the one hand, when the measurement devices are kept stationary (e.g., in a table), the Wi-Fi data collected cannot generate a reliable Wi-Fi fingerprint of the location, which may result in a very low recognition rate. On the other hand, when the measurement devices are detected to be at a location but are slightly moved (i.e., carried by a user), the Wi-Fi data can be used to generate a Wi-Fi fingerprint of this location which may result in a higher recognition rate. The records of the motion state can be used to improve the reliability of the fingerprint database.

5.5 Summary

In this chapter, the influence of small-scale fading and device heterogeneity on RSS values, as well as on the performance (i.e., recognition rate) of the Wi-Fi fingerprinting localization system WHERE was investigated. The experimental investigation shows that the recognition rate of the fingerprinting system is degraded significantly when the fading effect and the heterogeneous device problem exist. The recognition rate decreases from 100% to 47.76% when heterogeneous devices are used in the training phase and in the positioning phase. Due to small-scale fading, the fingerprints of the measurement positions, even only one centimeter apart in space, are regarded as different locations. In such a scenario, the localization system cannot recognize the locations nearby. However, the investigation shows that the collection of Wi-Fi data collected over a small distance can be used to generate the fingerprint of the location. In such a scenario, the recognition can be improved to 92.13%. Therefore, I propose to record the motion state of the measurement device when the training data is collected. In the real scenarios, the training data collected when the measurement devices are slightly moved (e.g., carried by a user) is more reliable than that the measurement device are kept stationary (e.g., on a table).

6 Conclusions

In this thesis, the influence of various factors including the fluctuation of Wi-Fi signal in an indoor environment, the use of different frequency bands, the use of heterogeneity devices, and the small-scale fading on the Wi-Fi signal and the performance of a Wi-Fi fingerprinting system were investigated. In chapter 2, I presented the state of the art of positioning applications and Wi-fi fingerprinting. I have clearly explained the principles of a Wi-Fi fingerprinting system, the challenges in implementing the system, the knowledge related to the development of Wi-Fi technology from the first to the up to date standard, the differences among different standards, different frequency bands, and new technologies that are implemented in new 802.11 standards. The studies from previous publications related to this thesis are surveyed and introduced, compared to each other. The methods to improve the performance of a Wi-Fi fingerprinting system are also examined.

In chapter 3, the fluctuation of the Wi-Fi signal in an office environment is investigated. The presence of people causes the higher fluctuation of Wi-Fi signal. The 5 GHz Wi-Fi signal has less fluctuation than the 2.4 GHz signal in a typical office environment. The degradation of 5 GHz signal through a wall is higher than the degradation of the 2.4 GHz signal.

In chapter 4, I compared the performance of a Wi-Fi fingerprinting system using 2.4 and 5 GHz signals. The results show that the performance of accuracy of Wi-Fi fingerprinting is similar when using 2.4 GHz and 5 GHz signals, but the recognition rate using signals from 5 GHz is higher than that of using signals from 2.4 GHz. Besides, scanning 2.4 GHz networks consumes less power than scanning 5 GHz networks.

In chapter 5, I investigated the influence of small-scale fading and device heterogeneity on RSS values, as well as on the performance of the Wi-Fi fingerprinting localization system. The recognition rate decreases from 100% to 47.76% when heterogeneous devices are used in the training phase and in the positioning phase. Due to small-scale fading, the fingerprints of the measurement positions, even only one centimeter apart, are regarded as different locations. However, the collection of Wi-Fi data collected over a short distance in space can be used to generate the fingerprint of the location. In such a scenario, the recognition can be improved to 92.13%.

The results of this thesis provide a better understanding of the different characteristics of the 2.4 and 5 GHz Wi-Fi signals. The 5 GHz signal tends to fluctuate in a smaller range and has a smaller fingerprint range than the 2.4 GHz signal. Therefore, the choice of the frequency band would have a significant impact on the performance of a Wi-Fi fingerprinting system. The accuracy of our fingerprinting system using the 5 GHz signal and the 2.4 GHz signal is similar. However, the recognition rate and the power consumption of the system using those two frequency bands are different. The findings recommend that to achieve a faster recognition rate, we should choose 5 GHz band for our Wi-Fi fingerprinting system. On the contrary, a fingerprinting system using the 2.4 GHz band helps to save power. The 5 GHz band has advantages over the 2.4 GHz band. More smartphones and APs are now equipped with both 2.4 and 5 GHz capability. Migrating to the 5 GHz band is a tendency of Wi-Fi networks.

The new techniques used in new 802.11 standards such as orthogonal frequency division multiplexing, channel state information, multiple antennas could provide more information about the transmission link between APs and measurement devices. This information can be utilized to mitigate the influence of the small-scale fading, multipath to the Wi-Fi signal strength as well as on the performance of a Wi-Fi fingerprinting system.

Today, smartphones are ubiquitous in many places. The number of mobile applications has been increasing rapidly in recent years. People use smartphones to do different kinds of tasks such as surfing the internet, checking Facebook, interacting with friends, and so on. Many applications take advantage of smartphones to provide location-based services. The physical locations are integrated into those services to offer users a better experience. Facebook and other popular social applications enable users to tag their photos, their posts with location information. People use location information not only on demand but also all the time to get updated information related to their physical locations. They use their smartphone to get information about their surroundings, use the information at their fingertips to learn more about the places they are interested in. Smart home, smart appliances, wearable devices, flying drone, and driverless car are a hot trend now and attract the attention from researchers. Many smart devices can connect to form a network and communicate with each other. Smartphones will play an important role as a hub to connect and control those smart devices. In order to interact with nearby smart devices, this model requires near constant location awareness. In the future, location awareness will not be considered as a separate function but will be incorporated into most network interactions, operating invisibly in the background.

A robust positioning system with high accuracy, the high recognition rate is a requirement for many applications and services. A positioning system could also combine various methods and techniques to improve its performance. The combination of Wi-Fi fingerprinting with inertial sensors, camera modules enable a highly accurate and robust positioning system. Proximity approach which uses sensors installed at reference places may be used to provide coarse location information. A Wi-Fi fingerprinting system incorporated with proximity approach to providing more specific, accurate location information. Location information from Wi-Fi fingerprinting may support other positioning methods available in a mobile device. GPS and Wi-Fi fingerprinting could be used together to locate the position both outdoor and indoor. Smartphones are ubiquitous and could be connected to form a dense mesh network. A specific smartphone could also get the location information of other nearby devices in the network to support for its positioning process. The direct device to device communication among smartphones enables for the application of cooperative positioning to improve the performance of location applications.

References

[1] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of Cellular Mobile Radio Localization Methods: From 1G to 5G," *IEEE Communications Surveys Tutorials*, vol. 20, no. 2, pp. 1124–1148, Dec. 2017.

[2] J. Frith, *Smartphones as Locative Media*. Polity Press, 2015.

[3] Y. Xu and K. David, "WHERE: An Autonomous Localization System with Optimized Size of the Fingerprint Database," in *International and Interdisciplinary Conference on Modeling and Using Context 2015*, Lanarca, Cyprus, 2015, pp. 544–550.

[4] Y. Xu, S. L. Lau, R. Kusber, and K. David, "DCCLA: Automatic Indoor Localization Using Unsupervised Wi-Fi Fingerprinting," in *Proceedings of International and Interdisciplinary Conference on Modeling and Using Context 2013*, Annecy, France, 2013, pp. 73–86.

[5] Y. Jiang *et al.*, "ARIEL: Automatic Wi-Fi Based Room Fingerprinting for Indoor Localization," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, Pittsburgh, PA, United States, 2012, pp. 441–450.

[6] C. Pinhanez, "Creating Ubiquitous Interactive Games Using Everywhere Displays Projectors," in *Proceedings of the International Workshop on Entertainment Computing (IWEC'02)*, Makuhari, Japan, 2002, pp. 149–156.

[7] J. R. Nogueras, "A Stream Redirection Architecture for Pervasive Computing Environments," Massachusetts Institute of Technology, 2001.

[8] N. Sukaviriya, M. Podlaseck, R. Kjeldsen, A. Levas, G. Pingali, and C. Pinhanez, "Embedding Interactions in a Retail Store Environment: The Design and Lessons Learned," in *Proceedings of the Ninth IFIP International Conference on Human-Computer Interaction*, Zurich, Switzerland, 2003, pp. 519–526.

[9] L. Chen, B. Li, K. Zhao, C. Rizos, and Z. Zheng, "An Improved Algorithm to Generate a Wi-Fi Fingerprint Database for Indoor Positioning," *Sensors*, vol. 13, no. 8, pp. 11085–11096, Aug. 2013.

[10] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of Wireless Indoor Positioning Techniques and Systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067–1080, Nov. 2007.

[11] R. Priwgharm and P. Chemtanomwong, "A Comparative Study on Indoor Localization based on RSSI Measurement in Wireless Sensor Network," in *2011 Eighth International*

Joint Conference on Computer Science and Software Engineering (JCSSE), Nakhon Pathom, Thailand, 2011, pp. 1–6.

[12] B. Bing, Ed., *Wireless Local Area Networks: The New Wireless Revolution*, First edition. Wiley-Interscience, 2002.

[13] F. Richard and A. Murat, *Location-Aware Applications*. Manning Publications, 2011.

[14] P. Bahl and V. N. Padmanabhan, “RADAR: an in-Building RF-based User Location and Tracking System,” in *Proceedings of IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, Tel Aviv, Israel, 2000, vol. 2, pp. 775–784.

[15] R. Want, A. Hopper, V. Falcão, and J. Gibbons, “The Active Badge Location System,” *ACM Transactions on Information Systems*, vol. 10, no. 1, pp. 91–102, Jan. 1992.

[16] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, “The Cricket Location-support System,” in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking*, Boston, Massachusetts, USA, 2000, pp. 32–43.

[17] N. B. Priyantha, A. K. L. Miu, H. Balakrishnan, and S. Teller, “The Cricket Compass for Context-aware Mobile Applications,” in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, Rome, Italy, 2001, pp. 1–14.

[18] S. Das, T. Teixeira, and S. F. Hasan, “Research Issues Related to Trilateration and Fingerprinting Methods,” *International Journal of Research in Wireless Systems*, vol. 1, no. 1, 2012.

[19] N. Swangmuang and P. Krishnamurthy, “An Effective Location Fingerprint Model for Wireless Indoor Localization,” *Pervasive and Mobile Computing*, vol. 4, no. 6, pp. 836–850, Dec. 2008.

[20] S. He and S.-G. Chan, “Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons,” *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 466–490, Firstquarter 2016.

[21] J. Hightower, S. Consolvo, A. LaMarca, I. Smith, and J. Hughes, “Learning and Recognizing the Places We Go,” in *Proceedings of International Conference on Ubiquitous Computing*, 2005, pp. 159–176.

[22] M. Youssef and A. Agrawala, “The Horus WLAN Location Determination System,” in *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services*, Seattle, Washington, USA, 2005, pp. 205–218.

[23] K. Kaemarungsi and P. Krishnamurthy, “Properties of Indoor Received Signal Strength for WLAN Location Fingerprinting,” in *Proceedings of The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004.*, Boston, MA, USA, 2004, pp. 14–23.

[24] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A Density-based Algorithm for Discovering Clusters a Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, 1996, pp. 226–231.

[25] O. Dousse, J. Eberle, and M. Mertens, “Place Learning via Direct WiFi Fingerprint Clustering,” in *Proceedings of IEEE 13th International Conference on Mobile Data Management*, Bengaluru, Karnataka, India, 2012, pp. 282–287.

[26] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, “OPTICS: Ordering Points to Identify the Clustering Structure,” in *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, Philadelphia, Pennsylvania, USA, 1999, pp. 49–60.

[27] Y. Xu, S. L. Lau, R. Kusber, and K. David, “An Experimental Investigation of Indoor Localization by Unsupervised Wi-Fi Signal Clustering,” in *Proceedings of 2012 Future Network Mobile Summit*, Berlin, Germany, 2012, pp. 1–10.

[28] S. Shin, A. G. Forte, A. S. Rawat, and H. Schulzrinne, “Reducing MAC Layer Handoff Latency in IEEE 802.11 Wireless LANs,” in *Proceedings of the Second International Workshop on Mobility Management & Wireless Access Protocols*, Philadelphia, PA, USA, 2004, pp. 19–26.

[29] M. Gast, *802.11ac: a Survival Guide*, First edition. O’Reilly, 2013.

[30] “IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,” *IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012)*, pp. 1–3534, Dec. 2016.

[31] A. Farshad, J. Li, M. K. Marina, and F. J. Garcia, “A Microscopic Look at WiFi Fingerprinting for Indoor Mobile Phone Localization in Diverse Environments,” in *Proceedings of 2013 International Conference on Indoor Positioning and Indoor Navigation*, Montbeliard-Belfort, France, 2013, pp. 1–10.

[32] Z. Yang, C. Wu, and Y. Liu, "Locating in Fingerprint Space: Wireless Indoor Localization with Little Human Intervention," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking, MOBICOM*, Istanbul, Turkey, 2012, pp. 269–280.

[33] C. Wu, Z. Yang, Y. Liu, and W. Xi, "WILL: Wireless Indoor Localization without Site Survey," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 4, pp. 839–848, Apr. 2013.

[34] A. Narzullaev, Y. Park, K. Yoo, and J. Yu, "A Fast and Accurate Calibration Algorithm for Real-Time Locating Systems based on the Received Signal Strength Indication," *AEU - International Journal of Electronics and Communications*, vol. 65, no. 4, pp. 305–311, Apr. 2011.

[35] R. Kubota, S. Tagashira, Y. Arakawa, T. Kitasuka, and A. Fukuda, "Efficient Survey Database Construction Using Location Fingerprinting Interpolation," in *Proceedings of IEEE 27th International Conference on Advanced Information Networking and Applications (AINA)*, Barcelona, Spain, 2013, pp. 469–476.

[36] A. Barry, B. Fisher, and M. L. Chang, "A Long-Duration Study of User-Trained 802.11 Localization," in *Mobile Entity Localization and Tracking in GPS-less Environments*, Springer, 2009, pp. 197–212.

[37] P. Bolliger, "Robust Indoor Positioning through Adaptive Collaborative Labeling of Location Fingerprints," ETH Zurich, Switzerland, 2011.

[38] S. Teller, J. Battat, B. Charrow, and D. Curtis, "Organic Indoor Location Discovery," MIT CSAIL Technical Report MIT-CSAIL-TR-2008-075, 2008.

[39] J. Park *et al.*, "Growing an Organic Indoor Location System," in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, San Francisco, California, USA, 2010, pp. 271–284.

[40] J. Ledlie *et al.*, "Molé: A Scalable, User-Generated WiFi Positioning Engine," in *Proceedings of 2011 International Conference on Indoor Positioning and Indoor Navigation*, Guimaraes, Portugal, 2011, pp. 1–10.

[41] M. A. Youssef, A. Agrawala, A. U. Shankar, and S. H. Noh, "A Probabilistic Clustering-Based Indoor Location Determination System," Technical Reports from UMIACS, Apr. 2002.

[42] A. K. M. M. Hossain, Y. Jin, W. S. Soh, and H. N. Van, “SSD: A Robust RF Location Fingerprint Addressing Mobile Devices’ Heterogeneity,” *IEEE Transactions on Mobile Computing*, vol. 12, no. 1, pp. 65–77, Jan. 2013.

[43] F. Dong, Y. Chen, J. Liu, Q. Ning, and S. Piao, “A Calibration-free Localization Solution for Handling Signal Strength Variance,” in *Mobile Entity Localization and Tracking in GPS-less Environnments*, Springer, 2009, pp. 79–90.

[44] M. B. Kjærgaard and C. V. Munk, “Hyperbolic Location Fingerprinting: A Calibration-Free Solution for Handling Differences in Signal Strength (concise contribution),” in *Proceedings of 2008 Sixth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2008, pp. 110–116.

[45] C. H. Wang *et al.*, “Robust Wi-Fi Location Fingerprinting Against Device Diversity based on Spatial Mean Normalization,” in *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Kaohsiung, Taiwan, 2013, pp. 1–4.

[46] J. Machaj, P. Brida, and R. Piché, “Rank Based Fingerprinting Algorithm for Indoor Positioning,” in *Proceedings of 2011 International Conference on Indoor Positioning and Indoor Navigation*, Guimaraes, Portugal, 2011, pp. 1–6.

[47] V. Honkavirta, T. Perala, S. Ali-Loytty, and R. Piche, “A Comparative Survey of WLAN Location Fingerprinting Methods,” in *Proceeding of Navigation and Communication 2009 6th Workshop on Positioning*, Hannover, Germany, 2009, pp. 243–251.

[48] R. Battiti, T. L. Nhat, and A. Villani, “Location-Aware Computing: a Neural Network Model for Determining Location in Wireless LANs,” University of Trento, Italy, Technical Report DIT-02-0083, 2002.

[49] D. A. Tran and T. Nguyen, “Localization in Wireless Sensor Networks Based on Support Vector Machines,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 7, pp. 981–994, Jul. 2008.

[50] C. P. Schloter and H. K. Aghajan, “Wireless Symbolic Positioning using Support Vector Machines,” in *Proceedings of 2006 International conference on Wireless Communications and Mobile Computing*, Vancouver, British Columbia, 2006, pp. 1141–1146.

[51] P. Sharma, R. K. Chaurasiya, and A. Saxena, “Comparison Analysis between IEEE 802.11a/b/g/n,” *International Journal of Scientific and Engineering Research*, vol. 4, no. 5, pp. 988–993, May 2013.

[52] Cisco, “802.11ac: The Fifth Generation of Wi-Fi Technical White Paper,” *Cisco and/or its affiliates*, 2018.

[53] G. Lui, T. Gallagher, B. Li, A. G. Dempster, and C. Rizos, “Differences in RSSI Readings made by Different Wi-Fi Chipsets A limitation of WLAN Localization,” in *Proceedings of 2011 International Conference on Localization and GNSS (ICL-GNSS)*, 2011, pp. 53–57.

[54] A. H. Ali *et al.*, “The Comparison Study of RF Signal Strength between IEEE802.11b/g and IEEE802.11n,” in *2011 IEEE 3rd International Conference on Communication Software and Networks*, Xi’an, China, 2011, pp. 646–649.

[55] J. Talvitie, M. Renfors, and E. S. Lohan, “A Comparison of Received Signal Strength Statistics between 2.4 GHz and 5 GHz Bands for WLAN-Based Indoor Positioning,” in *Proceedings of 2015 IEEE Globecom Workshops*, San Diego, CA, USA, 2015, pp. 1–6.

[56] V. Moghtadaiee and A. G. Dempster, “WiFi Fingerprinting Signal Strength Error Modeling for Short Distances,” in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2012, pp. 1–6.

[57] R. Faragher and R. Harle, “Location Fingerprinting with Bluetooth Low Energy Beacons,” *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2418–2428, Nov. 2015.

[58] A. M. Ladd, K. E. Bekris, A. Rudys, G. Marceau, L. E. Kavraki, and D. S. Wallach, “Robotics-based Location Sensing Using Wireless Ethernet,” in *Proceedings of the 8th Annual International Conference on Mobile Computing and Networking*, Atlanta, Georgia, USA, 2002, pp. 227–238.

[59] K. Kaemarungsi, “Distribution of WLAN Received Signal Strength Indication for Indoor Location Determination,” in *Proceedings of 2006 1st International Symposium on Wireless Pervasive Computing*, Phuket, Thailand, 2006, pp. 1–6.

[60] V. Bahl and V. Padmanabhan, “Enhancements to the RADAR User Location and Tracking System,” *Microsoft Research*, Feb. 2000.

[61] T. S. Rappaport, *Wireless Communication Principles and Practice*. Dorling Kindersley, 2009.

[62] R. Bernhardt, “Macroscopic Diversity in Frequency Reuse Radio Systems,” *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 5, pp. 862–870, Jun. 1987.

[63] K. Kaemarungsi and P. Krishnamurthy, “Analysis of WLAN’s Received Signal Strength Indication for Indoor Location Fingerprinting,” *Pervasive and Mobile Computing*, vol. 8, no. 2, pp. 292–316, Apr. 2012.

[64] Y. Chapre, P. Mohapatra, S. Jha, and A. Seneviratne, “Received Signal Strength Indicator and its Analysis in a Typical WLAN System (short paper),” in *38th Annual IEEE Conference on Local Computer Networks*, Sydney, NSW, Australia, 2013, pp. 304–307.

[65] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, “Tool Release: Gathering 802.11N Traces with Channel State Information,” *SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53–53, Jan. 2011.

[66] X. Wang, L. Gao, S. Mao, and S. Pandey, “CSI-based Fingerprinting for Indoor Localization: A Deep Learning Approach,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, Jan. 2017.

[67] Y. Chapre, A. Ignjatovic, A. Seneviratne, and S. Jha, “CSI-MIMO: Indoor Wi-Fi fingerprinting system,” in *Proceedings of 39th Annual IEEE Conference on Local Computer Networks*, 2014, pp. 202–209.

[68] Z. Zhuang, K.-H. Kim, and J. P. Singh, “Improving Energy Efficiency of Location Sensing on Smartphones,” in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, San Francisco, California, USA, 2010, pp. 315–330.

[69] M. Kjaergaard, “Location-based Services on Mobile Phones: Minimizing Power Consumption,” *IEEE Pervasive Computing*, vol. 11, no. 1, pp. 67–73, Jan. 2012.

[70] A. Haeberlen, E. Flannery, A. M. Ladd, A. Rudys, D. S. Wallach, and L. E. Kavraki, “Practical Robust Localization over Large-scale 802.11 Wireless Networks,” in *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking*, 2004, pp. 70–84.

[71] J. g Park, D. Curtis, S. Teller, and J. Ledlie, “Implications of Device Diversity for Organic Localization,” in *Proceedings of 2011 IEEE INFOCOM*, Shanghai, China, 2011, pp. 3182–3190.

[72] A. Tsui, Y.-H. Chuang, and H.-H. Chu, “Unsupervised Learning for Solving RSS Hardware Variance Problem in WiFi Localization,” *Mobile Networks and Applications*, vol. 14, no. 5, pp. 677–691, Oct. 2009.

[73] M. B. Kjærgaard, “Indoor Location Fingerprinting with Heterogeneous Clients,” *Pervasive and Mobile Computing*, vol. 7, no. 1, pp. 31–43, Feb. 2011.

[74] H. Aly and M. Youssef, “Dejavu: An Accurate Energy-efficient Outdoor Localization System,” in *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Orlando, Florida, 2013, pp. 154–163.

[75] M. Abdellatif, A. Mtibaa, K. A. Harras, and M. Youssef, “GreenLoc: An Energy Efficient Architecture for WiFi-based Indoor Localization on Mobile Phones,” in *2013 IEEE International Conference on Communications (ICC)*, 2013, pp. 4425–4430.

[76] I. Shafer and M. L. Chang, “Movement Detection for Power-efficient Smartphone WLAN Localization,” in *Proceedings of the 13th ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, Bodrum, Turkey, 2010, pp. 81–90.

[77] Y. Xu, L. Meng, and K. David, “Unsupervised Indoor Localization with Motion Detection,” in *Modeling and Using Context*, Springer, 2015, pp. 132–143.

[78] N. Brouwers, M. Zuniga, and K. Langendoen, “Incremental Wi-Fi Scanning for Energy-Efficient Localization,” in *Proceedings of 2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Budapest, Hungary, 2014, pp. 156–162.

[79] S. Wagner, M. Handte, M. Zuniga, and P. J. Marrón, “Enhancing the Performance of Indoor Localization Using Multiple Steady Tags,” *Pervasive and Mobile Computing*, vol. 9, no. 3, pp. 392–405, Jun. 2013.

[80] C. Fuchs, N. Aschenbruck, P. Martini, and M. Wieneke, “Indoor Tracking for Mission Critical Scenarios: A Survey,” *Pervasive and Mobile Computing*, vol. 7, no. 1, pp. 1–15, Feb. 2011.

[81] H.-Y. Wang, V. W. Zheng, J. Zhao, and Q. Yang, “Indoor Localization in Multi-Floor Environments with Reduced Effort,” in *Proceedings of 2010 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2010, pp. 244–252.

[82] T. King, T. Haenselmann, S. Kopf, and W. Effelsberg, “Key Factors for Position Errors in 802.11-based Indoor Positioning Systems,” University of Mannheim, Technical Report TR-2007-003, 2003.

[83] J. Hightower and G. Borriello, “Location Systems for Ubiquitous Computing,” *Computer*, vol. 34, no. 8, pp. 57–66, Aug. 2001.

[84] M. Baldauf, S. Dustdar, and F. Rosenberg, “A Survey on Context-Aware Systems,” *Ad Hoc and Ubiquitous Computing*, vol. 2, no. 4, pp. 263–277, Jun. 2007.

[85] A. Harter and A. Hopper, "A Distributed Location System for the Active Office," *IEEE Network*, pp. 62–70, 1994.

[86] R. Faragher and A. Rice, "SwiftScan: Efficient Wi-Fi Scanning for Background Location-based Services," in *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Banff, AB, Canada, 2015, pp. 1–6.

[87] "How t-Tests Work: 1-sample, 2-sample, and Paired t-Tests," *Statistics By Jim*, 06-Apr-2017. [Online]. Available: http://statisticsbyjim.com/hypothesis-testing/t-tests-1-sample-2-sample-paired-t-tests/. [Accessed: 13-Mar-2018].

[88] "T Test (Student's T-Test): Definition and Examples - Statistics How To," 13-Mar-2018. [Online]. Available: http://www.statisticshowto.com/probability-and-statistics/t-test/. [Accessed: 13-Mar-2018].

[89] Statpoint Technologies, *Statgraphics Centurion XVII User Manual*. http://www.statgraphics.com.

[90] Minitab Blog, "Understanding t-Tests: t-values and t-distributions." [Online]. Available: http://blog.minitab.com/blog/adventures-in-statistics-2/understanding-t-tests-t-values-and-t-distributions. [Accessed: 17-Aug-2018].

[91] "MATLAB." [Online]. Available: http://www.mathworks.com.

[92] S. L. Lau, Y. Xu, and K. David, "Novel Indoor Localisation using an Unsupervised Wi-Fi signal Clustering Method," in *Proceedings of 2011 Future Network and Mobile Summit*, Warsaw, Poland, 2011, pp. 1–8.

[93] Y. Xu, R. Kusber, and K. David, "An Enhanced Density-Based Clustering Algorithm for the Autonomous Indoor Localization," in *Proceedings of 2013 International Conference on MOBILe Wireless MiddleWARE, Operating Systems and Applications (Mobilware)*, Bologna, 2013, pp. 39–44.

[94] P. Kaveh and K. Prashant, *Principles of Wireless Networks: A Unified Approach*. Prentice Hall, 2002.

[95] A. U. H. Sheikh, *Wireless Communications Theory and Techniques*. Kluwer Academic Publishers, 2004.

[96] Solwise, "Pointer on Using the 5GHz WiFi Bands." [Online]. Available: http://www.solwise.co.uk/downloads/files/intheuk5ghz.pdf.

[97] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[98] Digi International, "Indoor Path Loss." [Online]. Available: http://ftp1.digi.com/support/images/XST-AN005a-IndoorPathLoss.pdf.

[99] P. Bolliger, “Redpin - Adaptive, Zero-configuration Indoor Localization Through User Collaboration,” in *Proceedings of the First ACM International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments*, New York, NY, USA, 2008, pp. 55–60.

[100] D. H. Kim, J. Hightower, R. Govindan, and D. Estrin, “Discovering Semantically Meaningful Places from Pervasive RF-beacons,” in *Proceedings of the 11th International Conference on Ubiquitous Computing*, Orlando, Florida, USA, 2009, pp. 21–30.

[101] I. Koenig, A. Memon, and K. David, “Energy Consumption of the Sensors of Smartphones,” in *Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS 2013)*, Illmenau, Germany, 2013.

[102] Y. Xu, D. Duong, and K. David, “How Near Is Near: A Case Study of the Minimum Distance to Distinguish Neighbouring Places in Place Learning Using Wi-Fi Signals,” in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, China, 2016, pp. 1–5.

[103] E. Frank, M. A. Hall, and I. H. Witten, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2011.